Formation of transcription factor complexes during embryonic erythroid development

Xiao Yu

ISBN: 978-94-6203-457-0

The studies and works presented in this thesis were performed at the Department of Cell Biology at the Erasmus MC, Rotterdam, The Netherlands. The department is a member of the Medisch Genetisch Centrum Zuid-West Netherland (MGC).

The research described in this thesis was financially supported by European Commission, the FP7 grant nr. 242129,
with project acronym SyBoSS.

Cover Design: Xiao Yu (于 晓/晓)
Printed by Wöhrmann Print Service, Zutphen

Formation of transcription factor complexes during embryonic erythroid development

Vorming van transcriptiefactor complexen tijdens de embryonale erytroïde ontwikkeling

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof. Dr. H.G. Schmidt
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
woensdag 2 oktober 2013 om 9.30 uur

door

Xiao Yu

geboren te Beijing

Promotiecommissie

Promotor:	Prof.Dr. F.G. Grosveld
Co-Promotor:	Dr. Charlotte Andrieu-Soler
Overige leden:	Prof.Dr. J.N.J. Philipsen
	Prof.Dr. D.F.E. Huylebroeck
	Dr. Ir. N.J. Galjart

Sometimes you will never know
the true value of a moment until
it becomes a Memory.

To my father and friends

Table of Contents

Abbreviations

DNA	deoxyribonucleic acid
RNA	ribonucleic acid
dpf	days post fertilization
ES cells	embryonic stem cells
BL-CFCs	blast colony forming cell
AGM	aorta gonads mesonephros
HSC	hematopoietic stem cell
MEL cells	mouse erythroleukemia cells
PCR	polymerase chain reaction
PLA	proximity ligation assay
RCA	rolling circle amplification
PPI	protein-protein interaction
BirA	biotin protein ligase
NuRD	nucleosome remodeling and deacetylase
RNA Pol II	RNA Polymerase II
MS data	Mass Spectrometry data
PAF1-C	RNA polymerase associated factor 1 complex
CDC73	Cell division cycle 73
IP	immunoprecipitation
RT-qPCR	reverse transcription quantitative PCR
shRNA	short hairpin RNA

Scope of the thesis

Hematopoiesis is a classic model for the study of embryonic and adult stem cell differentiation. Erythropoiesis is the process of generating erythrocytes from hematopoietic stem cells (HSC). In Chapter1, we introduce the process of erythropoiesis and discuss proteins and protein complexes that are essential for this process. In order to identify the temporal and spatial formation of LDB1 protein complexes, in Chapter2, we applied a new technology, the proximity ligation assay (PLA), to detect the protein complexes in fetal liver cells during embryonic development and in *in vitro* ES cell differentiation. In Chapter 3, we found that CDC73, a component of RNA Polymerase II associated factor complex (PAF1), can bind to LMO2 which is a component of LDB1 complex, suggesting it may be a link between the transcription factor LDB1 complex and the RNA polymerase II machinery. In Chapter4, we conclude our current results and discuss the future direction on LDB1 complex research.

Chapter 1

Introduction

Hematopoietic stem cells: emergence and differentiation

The process of blood production is called hematopoiesis, which generates all types of blood cells such as erythrocytes, megakaryocytes, or lymphocytes. This process derives from a rare pool of hematopoietic stem cells (HSC). Approximately 10^{11} to 10^{12} new blood cells are produced daily in an adult person. In an adult human body, approximately 2.4 million red blood cells (erythrocytes) transporting O_2 and CO_2 bound to hemoglobin are produced per second and circulate with a lifetime of 100-120 days. This massive amount of cells is generated by stem and progenitor cells to maintain our daily activities. Hematopoiesis has been classically used as a model system to study the regulation of stem cells and the mechanism of their differentiation into mature daughter cells while keeping self-renewal capacity.

In clinical stem-cell transplantations, HSC number is a crucial factor for the success of treatment [2]. The development of strategies to expand the pool of hematopoietic stem cells (HSCs) is one of the big challenges in regenerative medicine. Understanding of the microenvironment of HSCs, also called the “niche”, and its interaction with HSCs is thought to be important to improve the therapeutic techniques. This may also allow improve the reprogramming of other somatic cells such as fibroblasts directly into HSCs.

During mouse embryonic development, hematopoiesis occurs in three distinctive waves. The first wave, regarded as primitive hematopoiesis in the yolk sac, generates single-lineage progenitor cells which will drive the expansion of nucleated red blood cell at embryonic day 7.5 (E7.5). These red blood cells, called primitive erythroid cells, are present within blood islands in the yolk sac. These cells express embryonic hemoglobin [3]. The second wave happens in the aorta-gonad-mesonephros (AGM) region starting at E8.5, and gives rise to erythroid/myeloid progenitors (EMPs) between E8.5 and E10.5, which commit to primitive nucleated erythroid cells expressing adult hemoglobins and myeloid cells [4]. The third wave is the generation of adult long-term multi-lineage hematopoietic stem cells (HSCs) in the bone marrow, although the exact role of distinct HSC niches in the generation of HSCs is still arguable [5]. Currently, using live-imaging technique, these cells have been clearly showed to be derived from endothelia of AGM region at E10.5 [6]. Later these cells migrate into the fetal liver and differentiate into all types of blood cells. Following a short residence in fetal spleen, they eventually migrate to the bone marrow and stay there for the whole lifetime of the organism.

Due to the synchronous temporal and spatial emergence of endothelial and blood cells, the common precursor for these two lineages has been called hemangioblast [7]. This type of early precursor is present in low numbers in embryos which impedes its functional characterization *in vivo*. An *in vitro* ES cell-based system has therefore been developed to allow study of the earliest stages of hematopoietic commitment. Using this system, an endothelial-hematopoietic progenitor has been identified and

extensively been studied (**Figure 1**, [7-9]).

Since hematopoietic precursors emerge in distinct sites and at distinct developmental stages during embryogenesis, it is likely that they rely on distinct signaling mechanisms for their emergence. It has been shown that BMP, WNT, Hedgehog and NOTCH1 signaling pathways are essential for hemangioblast commitment and further development into primitive or definitive erythroid cells (**Figure 2**, [10]). *Bmp4* deficient mice die at E7.5-E9.5 with defects in mesoderm formation [11], which signaling activates expression of the genes encoding the *Gata2* and *Runx1* transcription factors, which are important for hematopoietic commitment [12, 13]. Along with extracellular signaling, intrinsic transcription factors also play key roles in cell fate determination and lineage commitment. This aspect will be described within the next section (« erythropoiesis »). The molecular regulations of ES cell differentiation into distinct stages from hemangioblast, hemogenic endothelium stage [14] to further blood cell lineages is summarized in **Figure 2**.

The fetal liver niche for HSCs contains several cell types including hematopoietic cells, hepatoblasts, macrophages, sinusoid endothelial cells and also mesenchymal cells [15]. It has been reported that DLK-1 expressing hepatoblasts express erythropoietin (EPO) and SCF to regulate erythroid differentiation [16]. Additionally, hepatoblasts produce angiopoietin-like 3, insulin-like growth factor-2, and thrombopoietin to increase the number of HSCs [16, 17].

In the bone marrow containing the adult HSC niche, the locations of HSCs are preferentially in close contact with SNO cells within bone marrow next to the trabecular bone to maintain the HSCs [18]. Another group identified that HSCs are close to bone marrow sinusoids near reticular cells expressing CXCL12/SDF-1 [19]. In addition, a model of hierarchical organization of mesenchymal and hematopoietic cross-talk in the bone marrow was proposed [20]. These findings suggest that self-renewal and differentiation properties of HSCs are controlled by extrinsic factors from the surrounding environment. The details of the effects of cytokines on HSC characteristics and regulators from the osteoblastic niche and vascular niches have recently been reviewed [21, 22].

In summary, hematopoiesis occurs at distinct sites of the embryo at different time points and hematopoietic development is influenced by distinct niches.

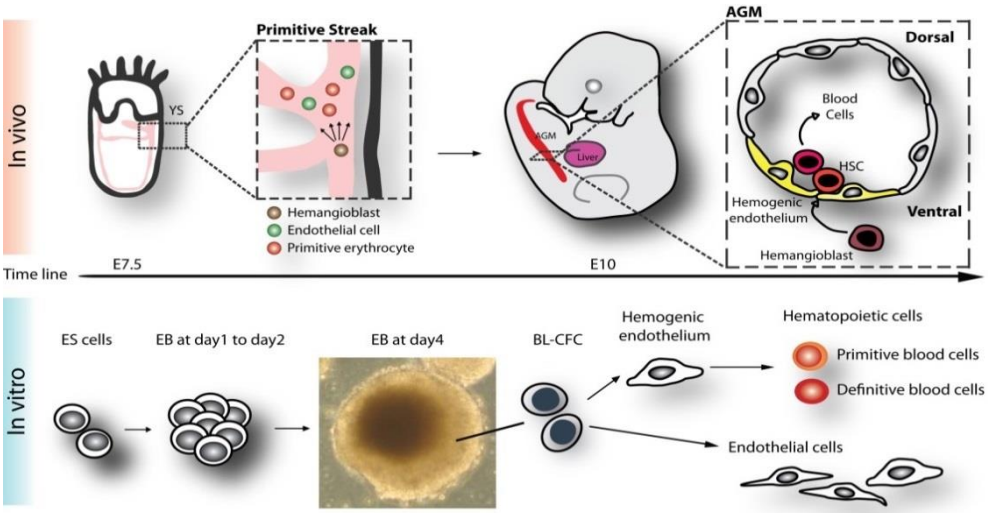


Figure 1. *In vivo* and *in vitro* model of hemangioblast differentiation

In the early stages of development (E7.5), hemangioblasts are transiently observed in the primitive streak in vivo before they migrate to the YS. Segregation of hematopoietic and endothelial lineages from hemangioblasts occurs before these progenitors reach their final destination in the YS. At later stages (E10), hematopoietic clusters appear in close association with the ventral aortic hemogenic endothelial cells in the AGM region (top panel). In vitro differentiation of ESCs involves generation of EBs and emergence of BL-CFCs, the in vitro equivalent of hemangioblast. BL-CFCs give rise to endothelial and hematopoietic cells. The latter differentiate through an intermediate hemogenic endothelium stage.

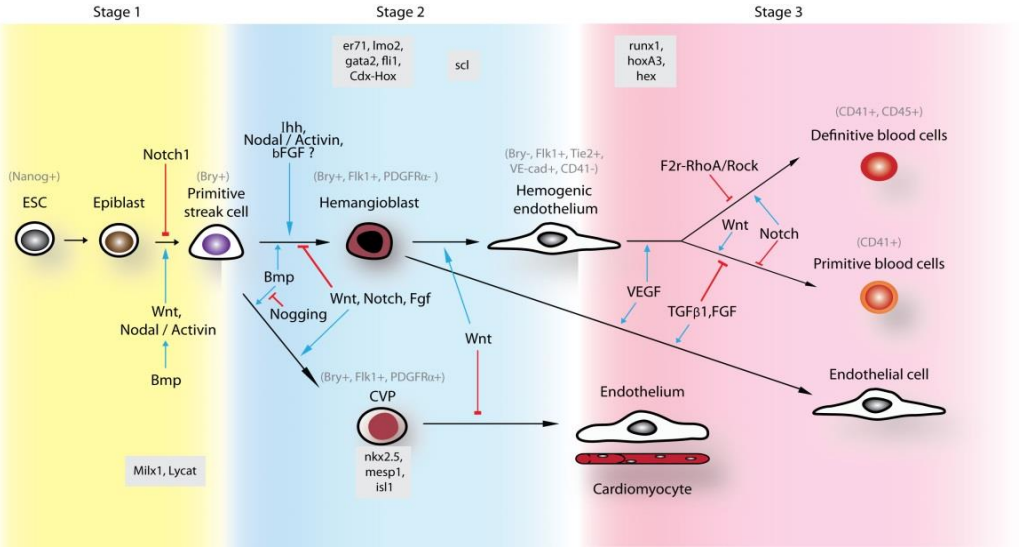


Figure 2. Schematic representation of hemangioblast/hemogenic endothelium emergence and differentiation.

Schematic figure describing the molecular regulation of ESCs differentiation into primitive streak-like cells (Stage 1),the competitive induction of hemangioblast versus CVP from mesodermal progenitor cells and the hemogenic endothelium emergence (Stage 2), and the later stages of hemangioblast/hemogenic endothelium differentiation towards endothelial and hematopoietic lineages (Stage 3). Markers for the identification of cell populations (grey letters), key TF regulators (grey boxes), and signaling pathways (blue or red arrows), are indicated. ESC, embryonic stem cells; CVP, cardiovascular progenitor cell; TF, transcription factor

(Figure1 and 2 are from a publication, Ref [1])

Erythropoiesis

Erythroblastic island

Erythropoiesis is the process of red blood cell formation. Generally, 4 to 5 cell differentiation stages are distinguished, from immature proerythroblasts, through basophilic erythroblasts, polychromatic erythroblasts, orthochromatic erythroblasts, reticulocytes and to fully differentiated erythrocytes (**Figure 3**, [23]). Cell sorting based on the membrane markers CD71 (transferrin receptor protein 1) and TER119 (Ly76, lymphocyte antigen 76) can distinguish the different stages of erythroid cell differentiation. CD71 is highly expressed in the early stages and decreases during differentiation; while TER119 is expressed at later stages. Erythropoiesis occurs in the erythroblastic islands (illustrated in **Figure 4-a**), which were discovered in 1958 by a French hematologist, Marcel Bessis. Since then, this erythroid-specific niche has been studied for positive and negative signaling that regulates the maintenance and differentiation of erythroid progenitors.

Erythroblastic islands contain a central macrophage serving as “nurse” cell to support erythroid differentiation and phagocytosis of extruded nuclei from the terminal stage of erythroblasts (**Figure 4-b** [24]). In bone marrow, the erythroblasts in one erythroblastic island are at various stages of differentiation. However, in the hypertransfused rat, erythroblasts become synchronized in each island [25]. It has been reported that the central macrophages harvested from human bone marrow express CD11a/c and CD18; while the counterparts in mice express F4/80 and Forssman glycosphingolipid [26]. Based on specific cell membrane markers, erythroblastic island components can be distinguished and reconstituted *in vitro* using TER119 for erythroblasts combined with F4/80 for macrophages (**Figure 4-c** [27]).

***In vitro* models of erythropoiesis**

Murine erythroleukemia (MEL) cells are first developed in 1971 by Friend et al.[28] and have been applied widely to study erythroid differentiation. These cells are derived from fetal liver erythroid progenitors and were transformed by Friend virus, leading to erythroleukemic cells blocked in their differentiation. Upon various chemical stresses, such as di-methyl sulfoxide (DMSO) treatment, these cells are able to differentiate and hemoglobinize *in vitro*, although only a small portion of cells enucleate [29]. These cells were therefore used to study the molecular mechanisms controlling the proliferation and differentiation of erythroid progenitors. Another widely used model, the human K562 cell line, was derived from a human chronic myeloid leukemia (CML) patient and is a model for human erythropoiesis [30]. G1E and G1E-ER4 cells are well studied to dissect the function of the GATA1 transcription factor. The G1E is a mouse GATA1-null cell line which is arrested at the erythroid progenitor stage. It was derived from GATA1-KO embryonic stem (ES) cells [31]. Rescue

by an estradiol-activated form of GATA1 allowed them to undergo terminal erythroid differentiation [32]. I/11 cells are also a model of erythroid progenitors. They were derived from *p53* knockout mouse fetal liver [29]. *In vitro* differentiation of wild-type embryonic stem cells can also give rise to ES-derived erythroid progenitors (ESEP) [33, 34].

Proteins involved in the erythroblastic island maintenance and function

In the erythroblastic island niche, cell-cell adhesive interactions and signals from environment are essential for erythroid development. Several cell-cell interaction signals have been discovered. For instance, erythroblast macrophage protein (EMP) mediates the attachment of erythroblasts and macrophages via its homophilic binding [35]. Inhibition of EMP/EMP interaction leads to decreased proliferation and increased apoptosis of erythroblasts [36]. VCAM-1/$\alpha_4\beta_1$ binding or ICAM-4/α_v integrin binding also mediates interactions of erythroblasts and macrophages. Inhibition of these interactions leads to breakdown of erythroblastic islands [27, 37].

The central macrophage functions in phagocytosis of nuclei from erythroblasts and transfer of iron for heme synthesis into the erythroid cells [38]. In a recent paper, Andrew et al. described the essential functions of CD169+ macrophages in both red-blood-cell production and clearance *in vivo* [39]. They showed that depletion of CD169+ macrophages reduced the number of erythroblasts in the bone marrow. Romas et al. also independently discovered that macrophage-erythroblast adhesion interactions are crucial in the erythroblastic expansion in induced anemia [40]. Proteins such as retinoblastoma tumor suppressor (Rb) protein and palladin are important for normal macrophage development and indirectly for red cell differentiation. Rb knockout mice are embryonic lethal with anemia due to abnormal enucleation [41, 42]. The cytoskeleton-associated protein palladin is essential for embryonic development. Homozygous palladin-null embryos die at E15.5 displaying defects including neural tube closure failure and severe anemia [43, 44].

A balance of positive and negative regulators in the niche is needed to maintain normal erythropoiesis. For example, physiological concentration of EPO positively regulates the number of erythroid progenitors preventing them to go into apoptosis [45]; while death receptor FAS/FASL binding mediates the number of red cells reaching maturity, functioning as negative regulator of fetal erythropoiesis [46]. Both tumor necrosis factor-α (TNF-α) and transforming growth factor-β (TGF-β) can inhibit immature erythroid cell proliferation [47-49]. Compared to steady state erythropoiesis, stress erythropoiesis is highly regulated by stress-induced factors including glucocorticoids [50], SCF [51], BMP4 [52, 53] and high level of EPO signaling [54].

In summary, the erythroblastic island is a structure forming a specialized microenvironment to regulate erythropoiesis, involving different mechanisms including cross-talk between erythroblasts and macrophages.

Transcription regulators in erythropoiesis

Cells need to read out the information stored in the DNA to produce functional proteins for homeostasis and development. A multitude of protein complexes is involved in transcription, such as the basal transcription machinery, transcription factors, mediator complex, chromatin remodeling complexes and histone (de)acetylases. In the next section, I focus on the basal transcription machinery and its associated complexes, and then on specific transcription factors such as GATA1 and LDB1 which are important for erythropoiesis.

The RNA Pol II complex and its regulation

Transcription, which generates RNA molecules from DNA, is one of the first and most important steps to regulate gene expression. The basic RNA polymerase II (RNA Pol II) complex, which initiates transcription at transcription start site (TSS) is a pre-requisite for all protein-coding genes and many non-coding RNAs. The RNA Pol II complex needs to be recruited at the promoter of the gene, to initiate RNA synthesis at its TSS. Following transcription initiation, elongating transcription factors stabilize the RNA Pol II complex to make it competent to generate full-length RNA transcripts.

Eight steps can be distinguished in the transcription cycle [55]. Each step is tightly regulated. Here I will focus on the step from transcriptional initiation to transcriptional elongation which is called “pausing”. It has been reported that RNA Pol II proximal promoter pausing is widely spread in the regulation of many genes [56, 57]. The transcription elongation factor b (P-TEFb) is a critical positive factor to switch the RNA Pol II from a pausing state to an actively elongating state in most genes [58, 59] through phosphorylation of Serine 2 (Ser2) of the carboxy-terminal domain (CTD) repeats on the RNA pol II [60]. A cyclin-dependent kinase (CDK9) and cyclin subunits (cyclin T1 and T2) were identified as subunits of P-TEFb in human [61]. Inhibition of P-TEFb by small compounds such as DRB or flavopiridol, leads to dramatic and global inhibition of transcription elongation by RNA Pol II [58]. DRB sensitivity inducing factor (DSIF) and the negative elongation factor (NELF) were identified to negatively regulate the elongation process by repressing Pol II elongation [62-64]. P-TEFb has been shown to phosphorylate DSIF to release NELF at pausing sites, leading to the initiation of elongation [65].

The RNA Pol II CTD contains tandem hepta-peptide repeats of consensus sequence (Tyr^1-Ser^2-Pro^3-Thr^4-Ser^5-Pro^6-Ser^7) that can be phosphorylated on Ser5 and Ser2 [66]. Phosphorylation on Ser5 is achieved by the CDK7 subunit of TFIIH, which primarily marks RNA Pol II that is located on the promoter region of genes and its phosphorylation level decreases during elongation [67]. In contrast, Ser2 phosphorylation is detected only on RNA Pol II that is actively transcribing and that is located in transcribed regions and is mediated by the CDK9 subunit of P-TEFb [68, 69].

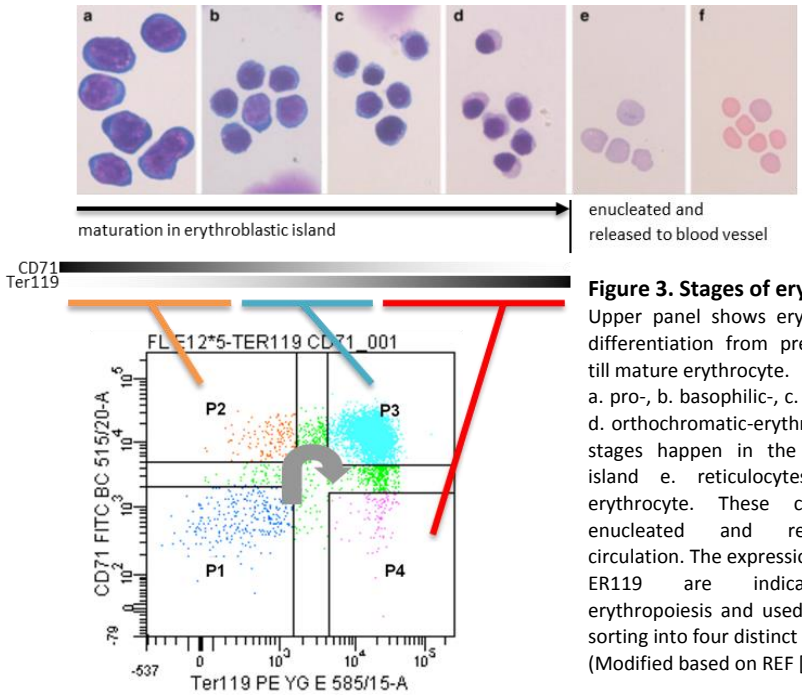
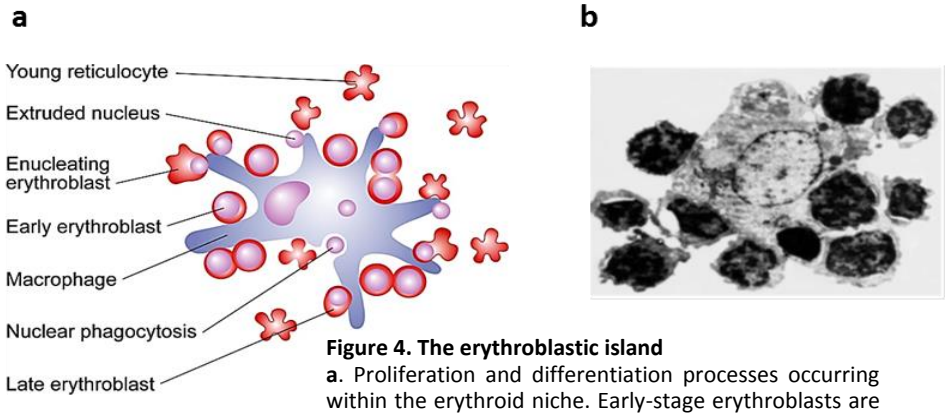


Figure 3. Stages of erythropoiesis.
Upper panel shows erythroblast cells differentiation from premature stage till mature erythrocyte.
a. pro-, b. basophilic-, c. polychromati-, d. orthochromatic-erythroblasts. These stages happen in the erythroblastic island e. reticulocytes, f. mature erythrocyte. These cells will be enucleated and released into circulation. The expression of CD71 and ER119 are indicated during erythropoiesis and used for fetal liver sorting into four distinct stages.
(Modified based on REF [22])



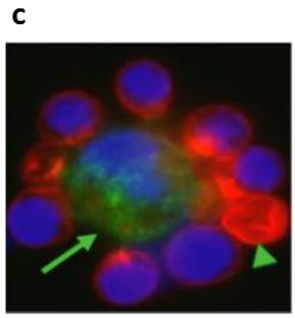


Figure 4. The erythroblastic island
a. Proliferation and differentiation processes occurring within the erythroid niche. Early-stage erythroblasts are larger cells with centrally located nuclei; more differentiated erythroblasts are smaller cells containing nuclei located adjacent to plasma membranes. Expelled nuclei undergo phagocytosis by central macrophage. Yong multilobulated reticulocytes are initially attached to the macrophage surface and later detach. Illustration by Paulette Dennis. **b**. Transmission electron micrograph of an erythroblastic island isolated from rat bone marrow. **c**. Confocal immunofluorescence image of an island reconstituted from freshly harvested mouse bone marrow cells stained with an erythroid-specific marker (red), macrophage marker (green) and DNA dye (blue). The central macrophage is indicated by an arrow and a multilobulated reticulocyte by an arrowhead. (These panels are from Reference [23].)

It has been reported that the phosphorylation level of Ser2 and Ser5 is a signal for 5' end capping and 3' end cleavage and polyadenylation events on the new transcripts [70, 71] and also triggers the interaction with splicing factors [72, 73]. In conclusion, a model of RNA Pol II transcription, involving Ser2 and Ser5 phosphorylation, DSIF and NELF association and capping/splicing events has been proposed [74]. In a recent paper, Ghamari et al. showed that CDK9 co-localizes with the Ser5 phosphorylated form of RNA Pol II which agrees with the role of the P-TEFb complex in regulating the early elongating RNA Pol II complex [75].

The Paf1 complex

Proteins interacting with RNA Pol II can influence polymerase activity to regulate transcription at different stages of the transcription cycle. One such complex, the Polymerase Associated Factor 1 (PAF1) complex (PAF1-C) of RNA Pol II mediators was isolated from a screen of yeast RNA Pol II associated proteins [76]. Its component cell division cycle 73 (CDC73), a gene product at *HRPT2* locus called parafibromin in humans, has been shown to directly interact with RNA Pol II *in vitro* [77, 78]. Later, all PAF1-C components were identified including PAF1, CDC73, CTR9, LEO1 and RTF1 in yeast [79]. This highly conserved complex also exists in Drosophila and humans [80, 81]. PAF1-C in humans also contains a SKI8/WDR61 protein [81]. In yeast, *Paf1* deletion dramatically decreases PAF1, LEO1, RTF1, CDC73 and CTR9 levels [82]. In zebrafish, RTF1 deficiency causes somite segmentation defects and morpholino knockdown embryos of *Rtf1* and *Ctr9* showed abnormal development of the heart and neural crest cells during embryogenesis [83-85]. In mice, homozygous deletion of *Cdc73* gene causes embryonic lethality at E6.5 and conditional deletion in adult leads to severe cachexia and death within 20 days [86]. In conclusion, PAF1-C is essential for embryonic development.

PAF1-C can influence RNA Pol II elongation by binding or recruiting various complexes. It has been reported that the subunits of the DSIF complex, Spt4p and Spt5p, are required for PAF1-C binding to RNA Pol II. PAF1-C interacts with "facilitates chromatin transcription/transactions" (FACT) complex to promote transcription elongation at actively transcribed genes [87]. The C-terminal domain of CDC73 contains a RAS family of small GTPase-like domain that facilitates full PAF1-C recruitment to activate genes [88]. This binding is in a phosphorylation-dependent manner which needs SPT5 and BUR1 (another elongation factor) to phosphorylate Ser2, 5 and 7 on RNA Pol II CTD. The phosphorylated SPT5 C-terminal repeats can also recruit PAF1-C [89].

It has been shown that PAF1-C can regulate transcription by promoting methylation of histone tails. In *Paf1* or *Rtf1* deleted yeast cells, H3 lysine4 (K4) methylation by COMPASS and DOT1p decreases, linking elongation and histone modifications [90]. Other histone modifications such as H3K36 tri-methylation by SET2 methyltransferase

[91] and H2B K123 mono-ubiquitylation [92, 93] also rely on functional PAF1-C. Accordingly, using chromatin IP (ChIP) on the *GAL7* upstream activation sequence, it has been shown that the amino acids 62-152 of yeast RTF1 have been identified as a histone modification domain (HMD) which can promote H3K4 and H3K79 methylation and H2BK123 ubiquitylation in yeast [94]. Deletion of PAF1-C components results in decreased mRNA polyadenylation due to the fact that PAF1-C associates with the cleavage and polyadenylation specificity factors (CPSF) complex and the RNA processing complex CSTF [95, 96].

Beside its positive role in transcription elongation, PAF1-C can also negatively regulate stem cell differentiation depending on the cell context. For instance, PAF1-C has been shown to mediate expression of cell-cycle related genes such as *c-Myc* and *Cyclin D* [97, 98] and in ChIP assays CDC73 is located at the promoter of genes involved in cell growth and survival, like *H19*, *Igf1*, *Igf2*, *Hmga1* and *Hmga2* [86]. Correspondingly, PAF1-C has been found to bind promoters of essential pluripotent genes such as *Oct4*, *Nanog* and *Sox2*, to maintain ESC identity [99]. Depletion of PAF1-C decreases an active transcription mark H3K4 trimethylation on pluripotent gene promoters and also reduces a repressive mark H3K27 trimethylation on cell fate gene promoters, which eventually leads to ES cell differentiation [99]. Additionally, the expression of PAF1-C components showed a decrease during hematopoietic differentiation [100]. In summary, PAF1-C appears to function in the stem cells or precursor cells to maintain their stages from differentiation.

The many functions of PAF1-C are mediated by its components. It has been shown that PAF1-C can link different signaling pathways: PAF1-C binds to the Hedgehog pathway Gli/Ci transcription factors to regulate their target genes [101]; RTF1 deletion causes reduction in histone H3K4 trimethylation via E3 ligase BRE1 which is crucial in Notch signaling [102] and the dephosphorylated form of CDC73 can bind to β-catenin which is an essential factor in Wnt signaling and which regulates Wnt target genes [103, 104]. PAF1-C is also linked to disease. PAF1-C has been shown to interact with mixed lineage leukemia (MLL) protein and its leukemic fusion MLL-AF9 protein together with super elongation complex (SEC) to regulate HOX gene expression in leukemic cells [105]. Furthermore, mutations in the *Cdc73* gene have been associated with hyperparathyroidism-jaw tumor (HPT-JT) syndrome [106, 107].

Gene expression is not only regulated by RNA Pol II transcription on the promoter region of a gene, the binding of transcription factors on tissue-specific enhancer regions is also critical. Transcription factors such as GATA1/LDB1 complexes binding to the enhancer and promoter regions can provoke DNA looping which involves protein-protein interactions between these regions. One example of looping is found in the *β-globin* gene locus [108, 109]. One complex essential for looping is the mediator complex which probably links between the upstream enhancer occupied by transcription factors and the promoter region bound by elongation factors and the

basal transcription machinery [110]. Interestingly, RNA Pol II is also found at the enhancer region of the *β-globin* gene [111]. Therefore, the complexity and hierarchy of erythropoiesis is tightly regulated by specific transcription factors which are able to activate or repress lineage- and stage-specific gene expression programs. Here I focus on a few specific transcription factor complexes that are important for erythroid development and differentiation.

GATA1 complexes

GATA1 is a key transcription factor determining erythroid fate and differentiation. The first site of *Gata1* expression is in nascent blood islands of the yolk sac [112, 113]. GATA1 is expressed in erythroid precursors, eosinophils, megakaryocytes, and mast cells, and also in testis [114, 115]. In the absence of GATA1, the proerythroblasts undergo apoptosis; while its overexpression inhibits terminal differentiation [116, 117]. Therefore, the precise level of GATA1 is essential for normal erythropoiesis. [118]. This property also applies to other essential transcription factors (TFs) such as *Myb.* Again precise levels of this TF are required to drive normal differentiation [119, 120]. The *Gata1* mutant in zebrafish is called *vlad tepes* and shows a severe reduction in blood cell precursors [121]. Mouse mutants are embryonic lethal at embryonic day 12.5 with severe anemia [118]. GATA1 is one of the GATA family members which all contain two highly conserved zinc finger domains that bind to an (A/T)GATA(A/G) consensus DNA motif in the genome. GATA2 is expressed in hematopoietic stem and progenitor cells (and many other tissues); while GATA3 is also expressed in HSCs [122], T-lymphocytes[123] and a number of other tissues, i.e. mammary gland[124]. Both GATA2 and GATA3 are also expressed in non-hematopoietic tissues such as the central nervous system [125] and during optic tectum [126]. *Gata1* knockout mice are embryonic lethal at E10.5-E11.5 with severe anemia [116]. Concordantly, in *Gata1* knockout ES cells driven to differentiate into red blood cells, a failure to generate mature red blood cells is observed, due to an inability to differentiate beyond the proerythroblastic stage [127]. Similar observations were made in GATA1 overexpressing cells [117]. Interestingly, ectopic GATA1 expression in lymphoid and granulocyte/monocyte (GM) progenitors reprogrammed these cells into megakaryocyte and erythrocyte lineages [128].

The essential GATA1 functions are accomplished through a complex protein-protein interaction network. A proteomics study of the GATA1 interactome carried out in MEL cells has revealed at least five different GATA1-containing complexes which may execute distinct functions (Fig5 [129]). GATA1 contains one transcriptional activation domain and two zinc finger domains regulating DNA binding. The C-terminal zinc finger (C-ZnF) is crucial for DNA binding. GATA1 binding partners interact with one of these two zinc fingers to form complexes. The FOG1/MeCP1 complex binds to the N-terminal ZnF of GATA1 to repress differentiation into other lineages and the

expression of specific progenitor genes. This complex is specifically recruited to the *Gata2* locus to repress its expression [129]. Recently, in a study of a naturally occurring mutant (GATA1-V205G) that cannot interact with FOG1, genome occupancy of wild-type and mutant GATA1 showed distinct binding sites which suggests that the non-DNA binding transcription factor FOG1 can modulate GATA1 binding [130]. The GATA1/GFI-1b complex represses genes related to cell proliferation, for instance *Myb* and *Myc*, to accomplish terminal erythroid differentiation [129]. The GATA1/TAL1/LDB1 complex primarily activates erythroid-specific genes. For example, the *Klf1* gene is specifically regulated by the binding of GATA1/TAL1 in its enhancer region but not by the GATA1/FOG1 complex [129]; meanwhile, disruption of the GATA1/FOG1 interaction decreases membrane protein expression, such as *Slc4a1*, *Spna1*, and *Aqp1* genes (encoding the membrane proteins band-3, α-spectrin, and aquaporin-1, respectively), but not of other GATA1-target genes such as *Alas2* [131]. The GATA1/TAL1 complex also activates well-known erythroid genes such as *glycophorin A* (*Gypa*) and the globin genes in MEL cells [132-134] and in human erythroid progenitors [135, 136]. The function of a GATA1 complex with ACF/WCRF/SNF2h is still unknown.

GATA factor switching

In the early stage of erythropoiesis, GATA2 is expressed and binds to its own promoter functioning as an activator [137]. Meanwhile, GATA2 can also bind to the *Gata1* gene to activate its expression. Once GATA1 expression is initiated, it represses *Gata2* by binding to its promoter with FOG1/MeCP1, while it continues to activate its own expression. This GATA factor change during erythropoiesis is called “GATA factor switching”, provoking changes in gene expression patterns during erythroid development [138, 139]. Thus, GATA2 regulates genes important for the proliferation of stem or progenitor cells and then during differentiation, GATA1 was proposed to determine erythroid fate and regulate the expression of erythroid-specific genes. However, it has been shown that spatio-temporal regulation of GATA factors is more important than their identity, since expression of *Gata1*, *2* and *3* under *Gata1* regulatory elements rescue the Gata1-null phenotype [140].

LDB1 complexes

LIM domain binding protein 1 (LDB1/CLIM2/NLI) together with GATA1 can form a multiprotein complex containing other transcription factors such as TAL1, LMO2 and RUNX1 to regulate erythroid development. LDB1 contains one homodimerization domain regulating long range interactions [141, 142]; one LDB1/Chip conserved domain (LCCD) that can bind SSBP proteins; a nuclear localization signal (NLS) domain; and importantly a LIM interaction domain (LID) which facilitates the specific interaction with LMO proteins such as LMO2 and LMO4 [143]. LDB1 is unable to

directly bind DNA, but stabilizes transcription complexes as a bridge/scaffold protein for various factors regulating diverse cellular processes [144].

The LDB1 protein is ubiquitously expressed and *Ldb1* knockout mice die embryonically at E9.5 to E10 from multiple developmental defects including disruption of midbrain and anterior hindbrain development, and extensive apoptotic cell death in mesenchymal tissue and absence of hematopoiesis [145]. *In vitro Ldb1-/-* ES cell differentiation gives rise to embryoid bodies (EBs) with a pale appearance compared to WT EBs ([9] and Chapter2). *Ldb1-/-* zebrafish also display various defects including deformation of the body axis and lack of hematopoiesis [146].

In conclusion, LDB1 is a core component of transcription complexes that regulate transcription in many cell types during embryogenesis.

The LDB1 complex, comprising a core of LDB1/GATA1/TAL1/E2A/LMO2 regulates many key erythroid master genes such as *Tal1*, *Gata1*, *Klf1*. This complex also binds to the locus control region and *β-globin* promoter in murine erythroleukemic (MEL) cells [147]. LDB1 was also shown to participate in the control of a transcriptional program regulating HSC maintenance [148], but has a critical role in heart formation and anterior patterning of the gastrulating embryo [9, 145]. It was also shown to play important functions in the regulation of normal homeostasis of adult intestinal stem cells. Loss of LDB1 leads to the decrease of intestinal LGR5+ stem cells and increases cell proliferation in the epithelia through activation of WNT signaling, eventually triggering apoptosis of the organ [149]. Interestingly, LDB1 does not only act in the nucleus to regulate transcription, but also seems to have a role in the cytoplasm to mediate cell migration in fibroblastic cells through Ste20 kinase SLK [150].

TAL1 (also called SCL) belongs to the basic helix-loop-helix (bHLH) transcription factor family which regulates cell fate and differentiation [151]. It was originally identified in a human leukemic cell line [152]. *Tal1* knockout mice are embryonic lethal at E10.5 with impaired primitive hematopoiesis and no blood islands detected in their yolk sacs [153]. GATA1 and PU.1 are neither expressed in *Tal1* knockout mice nor in *Tal1-/-* ES cell differentiated EBs. This suggests that GATA1 and PU.1 are downstream of TAL in the hematopoietic regulatory network [154, 155].

LMO2 and LMO4 belong to the LIM-Only protein family which regulates protein-protein interactions through the LIM domains. LMO2 is important both in mice and in zebrafish hematopoiesis. *Lmo2* deletion in mice causes embryonic lethality at E10.5 with no blood islands [156]. Knockdown zebrafish shows a loss of primitive erythropoiesis [146, 157]. *Lmo4* knockout mice show no hematopoietic defects while they have neural tube closure defects [158, 159]. However, *Lmo4-/-* zebrafish are pale with decreased *Runx1* expression levels in the dorsal aorta region [146], indicative of impaired hematopoietic development.

RUNX1 is another transcription factor that regulates cell fate and differentiation. *Runx1*-/- mice die at E11.5 to E12.5 with severe hemorrhages; however, primitive hematopoiesis is not affected. Interestingly, in these mice definitive hematopoiesis is absent both in the AGM and fetal liver, displaying no erythroid/myeloid progenitors (EMPs) and long-term HSCs [160]. Consistent with this phenotype, RUNX1 is expressed in the ventral wall of the dorsal aorta just prior to the emergence of the HSCs [161] and provides important activity in maintaining normal HSC function [6] and the terminal differentiation of megakaryocytic maturation and lymphocytic differentiation [162]. RUNX1 has been shown to be a complex with LDB1, GATA1, TAL1 and ETO2 [146]. Recently, it has been found that RUNX1 can interact with LSD1 and MYEF2 in erythroid cells. The interaction of RUNX1/MYEF2 can repress its target genes such as *Eto2* and *Gata1* in undifferentiated MEL cells [163].

Signaling pathways in erythropoiesis

EPO-mediated signaling

Erythropoietin (EPO) secreted from the kidney is a crucial regulator for red blood cell production. It has been identified as one of survival signals of erythroid progenitors in the fetal liver for embryonic erythroid development and bone marrow and spleen in adult stage [164]. Its receptor EPO-R is expressed during the earliest stages of erythroid development i.e. in proerythroblasts [165]. Mice lacking EPO or EPO-R show embryonic lethality with severe anemia during embryogenesis at E12.5 to E13.5. However, burst-forming unit-erythroid (BFU-E) and colony-forming unit-erythroid (CFU-E) are detected, indicating that EPO/EPO-R signaling pathways are important for erythroid expansion and survival from these stages onwards [166].

The binding of EPO to the dimerized EPO-R leads to conformational changes in the cytoplasmic region of the EPO-R and activates Janus Kinase 2 (JAK2), Lck/Yes-related novel tyrosine kinase (LYN) and Tec family Bruton's tyrosine kinase (BTK) to phosphorylate several tyrosine sites on EPO-R. LYN-deficient cells have absent or decreased expression of *Gata1*, *Klf1* and *Stat5*, indicating that these genes are up-regulated downstream of EPO/EPO-R/LYN signaling [167]. BTK is also activated by EPO. Contrary to *Lyn*, *Btk*-deficient mice have an enhanced erythroid differentiation capacity [168].

Several signaling pathways are affected by EPO including a number of positive pathways: (i) a few STAT family members, STAT1, STAT3 and STAT5a/b, (ii) ERK1/2, SAPK/JNK and p38, (iii) PI 3-kinase/PKB/Akt/Foxo3a (i-iii, [169]); and some repressive mechanisms: (iv) SOCS [170], (v) SH2-domain containing phosphatase, SHP1 [171], SHP2 [172], CD45 [173] and PTB-1B [174].

In conclusion, the EPO/EPO-R signaling pathway is important to regulate erythroid expansion, survival and differentiation in embryonic, fetal and adult stages.

SCF and c-KIT signaling

Stem cell factor (SCF) receptor (CD117, known as c-KIT) is expressed in the immature hematopoietic compartment (including HSCs) and in the early stage of erythroid differentiation. Gain of function mutations of SCF/c-KIT signaling are related to a variety of malignancies [175]. Both c-KIT (a product of the *W* locus) and its ligand SCF (a product of the *Sl* locus) are essential in gametogenesis, melanogenesis and hematopoiesis [176-178]. *c-Kit*-/- mice die perinatally, showing severe macrocytic anemia which starts from embryonic day 11 (E11), while knockout of SCF leads to severe microcytic anemia after E13 [178, 179]. Consistently, mouse mutants with *lacZ* reporter gene under control of the promoter of *c-Kit* creating a null allele showed that c-KIT is expressed in the AGM region establishing the normal fetal hematopoietic precursors, which indicates that *c-Kit* is not required for the primitive hematopoiesis [180]. The phenotypic differences of the *Scf* and *c-Kit* knockout animals may indicate that primitive versus definitive HSCs self-renewal activity may be regulated differently in their normal distinct niches [181].

REDS

Red cell differentiation signal (REDS) has been proposed by Whyatt et al. in 2000 [117]. They found that *Gata1* overexpressing mice die from severe anemia due to a block in differentiation of these cells at the proerythroblast stage. Interestingly, these cells are able to be differentiated *in vivo* in the presence of wildtype cells in chimeric mice [117]. This suggests that a signal from wild-type cells can overcome the defect of *Gata1* overexpression specifically for erythroid differentiation. It was shown that REDS is expressed by the normal differentiating erythroid cells to regulate the balance between erythroid proliferation and differentiation in the erythroblastic island [182].

Hematopoiesis and erythropoiesis in zebrafish

In addition to humans and mice, a number of other model systems have been used to study the process of hematopoietic development, in particular Xenopus and more recently Zebrafish. Zebrafish (*Danio rerio*) have been widely used as a non-mammalian model for several reasons. Its similarity to mammals in the genetic programs regulating hematopoiesis [183, 184], its transparency allowing to directly observe internal organs and blood production, short embryonic developmental time and its amenability to large-scale screening, make it extremely suitable to dissect the mechanism of normal hematopoiesis.

The two waves of hematopoiesis, primitive and definitive, also arise during zebrafish development. Primitive hematopoiesis occurs within 30 hours postfertilization (hpf) which mimics the mouse and human embryonic hematopoiesis [185]. In zebrafish, primitive hematopoiesis develops intraembryonically from two different locations: (i) the ventral mesoderm pole, becoming caudal/posterior lateral plate mesoderm (LPM) expressing *gata1*, and contributing the intermediate cell mass (ICM) leading to the emergence of primitive erythroid cells [186]. (ii) the dorsal site in the blastula, the rostral/anterior LPM expressing *spi1* (also known as *pu.1*) but not *gata1*, is contributing to cardiac-fated cells and also exhibits behavior consistent with a macrophage identity [187]. Definitive hematopoiesis in zebrafish also mimics the mammalian counterpart. It has been shown that definitive hematopoiesis has two distinct phases. The first phase is the generation of erythroid/myeloid progenitors (EMPs) to generate erythroid and myeloid cells [188]. The second phase is the process of hematopoietic stem cell (HSC) differentiating to all blood cell types. The HSC-fated cells derive from the ventral wall of the dorsal aorta, which is the counterpart of aorta-gonad-mesonephros (AGM) region in mammals, where cells express the definitive marker *c-Myb* [189]. Later, these cells migrate to the posterior region in the tail, the caudal hematopoietic tissue (CHT) [190]. Eventually, the HSCs will reside in the kidney, as opposed to the bone marrow in mammals, and which becomes the main hematopoietic organ in adult zebrafish.

Transcription factors have been extensively studied in zebrafish using morpholino-mediated knockdown. This approach showed that *scl*/*tal1*, *gata1*, *ldb1*, *lmo2* and *gata2* are important factors for establishing normal hematopoietic development. For instance, absence of *scl* in zebrafish embryos results in the complete loss of primitive erythropoiesis and myelopoiesis and a lack of *c-myb* and *Runx1* expression in the dorsal aorta region [191]. *Gata1* knockdown embryos show anemia at 36 hpf and convert the cell fate in the ICM region into myeloid cells [192]. *ldb1* knockdown leads to defects in definitive hematopoiesis whereas *lmo2* knockdown results in a loss of primitive erythropoiesis [146, 157]. *Lmo4* knockout mice die embryonically due to a failure of neural tube closure but have normal hematopoiesis [159]; meanwhile, MO knockdown of *lmo4* in zebrafish is lethal with severe anemia

[146]. Interestingly, the absence of *Gata2* in zebrafish results in mild defects in primitive hematopoiesis, whereas *Gata2*-/- mouse embryos show embryonic lethality associated with severe anemia [192, 193].

In forward genetic screening in zebrafish, several hematopoietic mutant models have been identified and used to dissect the hematopoietic regulatory mechanisms. As an example, the *cloche* mutant was discovered in 1995 by Stainier [194] and in 2004 Davidson [184] found that the mutant has complete absence of hematopoietic and endothelial cells while other mesodermal tissues were normal. This phenotype indicates that this mutation affects the progenitors of both hematopoietic and endothelial cells, i.e. the hemangioblast. In agreement with the *gata1* knockdown previously mentioned, cells in the ICM originally designated to become erythrocytes switch their fate to myeloid cells. The *tif1-γ* gene mutant zebrafish is called *moonshine* for its excessive amount of pigment cells in the body [195] and shows defects in both primitive and definitive erythropoiesis; meanwhile myelopoiesis is normal [196]. Mouse embryos deficient in *Tif1-γ* gene show a developmental delay and die at embryonic day 9.5 [197]. In zebrafish, it has been shown that the *sunrise* mutant, which has a mutation in the *Cdc73* gene, can rescue the erythroid defect in the *moonshine* mutant (*Tif1-γ* deficient) animals [198]. Immunoprecipitation experiments showed that TIF1-γ binds to GATA1/TAL1/LDB1 complex, together with positive elongation factors P-TEFb and FACT, to regulate blood genes in K562 cells. TIF1-γ and CDC73 share common gene targets such as *Gata1* and *Gata2* [198].

As discussed previously, TFs are key components of the hematopoietic regulatory networks. TFs form multiprotein complexes, the composition of which ultimately determines positive or negative transcriptional activity. For instance GATA1 can form distinct complexes at different time points to regulate blood genes [129] and the negative regulators ETO2 and MTGR1 in the LDB1 complex decrease during differentiation [146]. This feature underscores the need to precisely define TF complex composition throughout development. Several techniques have been developed to allow precise measurement of protein interactions in different cellular contexts. Some of these are summarized below.

Protein-protein interaction techniques

Normal physiological cellular functions are mainly mediated through protein-protein interactions (PPIs), and the majority of proteins function as components of larger complexes that regulate specific biological functions. Hence, understanding PPIs is essential to understand on its turn the biological processes within a cell, including those resulting from cell-cell and cell-matrix communication. Different methods of PPI detection have been developed. Here, I will present only the widely used techniques, including recent ones.

Protein pull-down

The most frequently used techniques are *in vitro* pull-down assays for epitope-tagged proteins using anti-epitope antibody conjugated solid supports and *in vivo* co-immunoprecipitation (co-IP) with either anti-epitope antibody or specific antibodies targeting the protein(s) of interest. These assays may be coupled to mass spectrometry in order to identify novel directly interacting partners and/or other proteins of the complex without prior knowledge of their identity. Provided that high quality antibodies or affinity columns are available, these methods are in general very efficient although they still requires relatively large amounts of starting material (i.e. lysates of sufficient cells). In addition, the experimental conditions need to be controlled carefully in order to avoid non-specific interactions. For instance, the salt concentration can change the binding affinity of individual polypeptides in macromolecular complexes, explaining low amount and/or numbers of binding partners in high-salt condition. It can therefore also be difficult to detect transient protein-protein interactions using pull-downs.

Yeast two-hybrid

Another method for identifying PPIs is the yeast two-hybrid (Y2H) assay, originally developed by Stanley Fields [199]. The method is based on a trans-activation assay in yeast using a specific coding cDNA (or a part thereof) or a cDNA library fused to a transactivation domain (AD). These “preys” are introduced into yeast containing a reporter or selectable marker gene with binding sites for a well-characterized DNA binding (DB) domain (i.e. of GAL4) fused in-frame to the protein (i.e. transcription factor) which is used as a “bait”. The physical interaction of the two polypeptides of interest will bring the DB and AD in proximity, resulting in the activation of the reporter/selection genes. The Y2H method in mammalian cells has enabled to demonstrate TGF-β stimulation dependent association of SMAD2 and SMAD4 with CREB binding protein [200]. It has been shown that the strength of interaction detected in two-hybrid generally correlates with that determined in *in vitro* assays [201]. This “old” technique has been used widely and in diverse variants, and became amenable to high-throughput applications in mammalian cells as well, leading to the

large-scale detection of PPIs and networks, i.e. the interactomes [202]. cDNA library based screening approaches can provide large number of PPIs, including LUMIER (luminescence-based mammalian interactome mapping) [203], but the latter detects mostly non-specific interactions, hence this technique has also become less popular. In a recent research on the comparison of yeast two-hybrid systems, it has been shown that interaction detected in the two-hybrid approach generally correlates with *in vitro* experiments [204], nevertheless because it yielded many false positives and is much slower than modern proteomic approaches, it has gone out of fashion.

FRET/BRET

A fluorescence-based technique to detect PPI in living cells has been developed with nanometer resolution based on Förster resonance energy transfer (FRET) [205] or bioluminescence resonance energy transfer (BRET) [206]. In these, the proteins of interest are fused to distinct fluorophores, a donor fluorescent protein (FP) and an acceptor FP. When the interaction of the candidate-interacting proteins takes place, the two FPs are brought together within 2 and 8 nm [207]. By laser excitation of the donor FP, the resonant energy can transfer from donor to acceptor, thereby exciting the acceptor FP. The emission of the acceptor FP can be distinguished from the wavelength of the donor FP, and is indicative of an interaction between the two proteins. As an alternative to the external laser, which may increase the noise (FRET), BRET uses luciferase to produce the excitation energy. Since FRET and BRET are based on resonance energy transfer, the emission of the acceptor FP will be terminated when the two proteins are no longer binding to each other. Thus, both methods can monitor spatial-temporal changes of protein interactions of interest and can specifically detect transient interactions in real-time. However, due to the short distance needed between two FPs, they are limited in detecting PPIs in macromolecular complexes. In addition, the absence of a signal does not always mean absence of interaction. Meanwhile, these advanced techniques are claimed to depend significantly on (i) the transfection efficiency of constructs in a cell, (ii) the donor FP life-time and (iii) the overlap of donor emission and acceptor excitation spectra. As a result, a good combination of two optimal FPs is needed.

BiFC

Another fluorescence-based technique is bimolecular fluorescence complementation (BiFC). In BiFC, a fluorescent protein (i.e. GFP, YFP) is split into two specific fragments which by themselves cannot give rise to any fluorescence. These two split fragments are each fused to a distinct candidate protein. If these candidate proteins physically interact, the two split fragments are brought into sufficiently close proximity to recreate a fully functional fluorescent protein (i.e. GFP). Detection of a fluorescence signal is therefore indicative of PPIs. The pioneering BiFC experiment in

living mammalian cells was performed in 2002 using enhanced YFP-based BiFC; it uncovered the localization of interactions between the basic leucine zipper (bZIP) and NF-κB family proteins [208]. This study promoted a wider usage of BiFC. However, the method has several disadvantages: (i) The self-assembly of the non-fluorescent fragments can lead to a reduction of the signal-to-noise (S/N) ratio, (ii) The split site of the fluorophore is also critical for this S/N ratio and (iii) The BiFC technique is irreversible [209].

PLA

The proximity ligation assay (PLA) is based on the specificity of antibody and polymerase chain reaction (PCR) to detect PPI. Antibodies targeting proteins of interest are incubated on fixed cells or tissue with reagent like PFA. Secondary antibody conjugated with oligo-DNA probes are sequentially added to target specific species of primary antibodies. Followed by hybridization of oligo-DNA and ligation, PCR has been performed. After hybridization of detection probes on the amplified DNA, single molecule or PPI can be detected. In 2002, Ulf and co-workers succeeded in monitoring single-molecule dimerizations *in vitro* using this assay for platelet-derived growth factor B-chain (PDGF-BB) protein binding [210]. They also analyzed several growth factors and cytokines to compare the results from the PLA assay, enzyme-linked immunosorbent assay (ELISA) and Western blot assay [211] and concluded that the PLA assay can detect 100-fold lower protein concentration in comparison to ELISA, which is the most convenient method to detect low amounts of proteins in a sample.

In a new version of the PLA assay, oligonucleotide probes are directly attached to the proteins via bifunctional cross-linking reagents or conjugated to mono- or polyclonal antibodies that recognize the proteins [211]. By modifying the PLA protocol, the *in situ* proximity ligation assay (*in situ* PLA) [212] uses rolling-circle-amplification (RCA) by the enzyme Phi29 polymerase to obtain high sensitivity. (For the detailed PLA assay method and scheme, please see **Appendix**.) It replicates a 100-mer DNA circle 1000 times on-site in one hour at 37C^{o}. Adding detection probes conjugated to a fluorophore can generate a bundle (or blob) of DNA (less than 1μm in diameter), that can be observed under normal microscopes. Immuno-fluorescence (IF) can detect the co-localization of two or several proteins but only distinguishable in a distance of few-hundred nm; while PLA can detect at a much higher resolution down to 35 to 40nm. Furthermore, PLA can detect a single individual targeted protein leading to the possibility of quantitatively measuring single molecules in a cell. It is not limited to detect single molecule, as PLA assay can also detect subcellular localization in a solid-phase sandwich assay like IF. It can also detect the endogenous level of a target protein, a protein modification by using one specific antibody targeting the modification site, or using two antibodies targeting the proteins of interest for PPI.

Although the PLA assay has valuable advantages, it cannot detect real-time PPIs or transient PPIs in live cells like FRET/BRET.

Comparison of distinct PPI methods

Each technique mentioned above has its own unique merits and disadvantages. When choosing one method versus another, the purpose of the experiment and the technical feasibility should be the main concern.

The table below shows the merits for each PPI technique. For instance, if the PPI of interest happens very transiently and the researcher would like to visualize it; the best method would be FRET/BRET. If the endogenous level of a protein and its specific localization is most important for example in a patient sample, the choice would be PLA or IF.

	PLA	IF	IP	Y2H	FRET/BRET	BiFC
Detect single molecules	✓					
Detect *in situ*	✓	✓			✓	✓
Endogenous level	✓	✓	✓			
Protein-protein interaction (PPI)	✓		✓	✓	✓	✓
Detection in tissue	✓	✓		✓		
Quantitative	✓					
Live cell				✓	✓	✓
Reversable					✓	

References

1. Andrea Martella, X.Y., Charlotte Andrieu-Soler and Frank Grosveld, *Transcriptional regulation and signalling pathways involved in the hemangioblast development.* Trends in Developmental Biology, 2012. **6**: p. 53-73.
2. Grewal, S.S., et al., *Unrelated donor hematopoietic cell transplantation: marrow or umbilical cord blood?* Blood, 2003. **101**(11): p. 4233-44.
3. Palis, J. and M.C. Yoder, *Yolk-sac hematopoiesis: the first blood cells of mouse and man.* Exp Hematol, 2001. **29**(8): p. 927-36.
4. Dzierzak, E. and N.A. Speck, *Of lineage and legacy: the development of mammalian hematopoietic stem cells.* Nat Immunol, 2008. **9**(2): p. 129-36.
5. Medvinsky, A., S. Rybtsov, and S. Taoudi, *Embryonic origin of the adult hematopoietic system: advances and questions.* Development, 2011. **138**(6): p. 1017-31.
6. Chen, M.J., et al., *Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter.* Nature, 2009. **457**(7231): p. 887-91.
7. Lancrin, C., et al., *Blood cell generation from the hemangioblast.* J Mol Med (Berl), 2010. **88**(2): p. 167-72.
8. Choi, K., et al., *A common precursor for hematopoietic and endothelial cells.* Development, 1998. **125**(4): p. 725-32.
9. Mylona, A., et al., *Genome-wide analysis shows that Ldb1 controls essential hematopoietic genes/pathways in mouse early development and reveals novel players in hematopoiesis.* Blood, 2013. **121**(15): p. 2902-13.
10. Chiang, P.M. and P.C. Wong, *Differentiation of an embryonic stem cell to hemogenic endothelium by defined factors: essential role of bone morphogenetic protein 4.* Development, 2011. **138**(13): p. 2833-43.
11. Winnier, G., et al., *Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse.* Genes Dev, 1995. **9**(17): p. 2105-16.
12. Lugus, J.J., et al., *GATA2 functions at multiple steps in hemangioblast development and differentiation.* Development, 2007. **134**(2): p. 393-405.
13. Pimanda, J.E., et al., *The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 840-5.
14. Lancrin, C., et al., *The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage.* Nature, 2009. **457**(7231): p. 892-5.
15. Sugiyama, D., et al., *Embryonic regulation of the mouse hematopoietic niche.* ScientificWorldJournal, 2011. **11**: p. 1770-80.
16. Sugiyama, D., et al., *Hepatoblasts comprise a niche for fetal liver erythropoiesis through cytokine production.* Biochem Biophys Res Commun, 2011. **410**(2): p. 301-6.
17. Chou, S. and H.F. Lodish, *Fetal liver hepatic progenitors are supportive stromal cells for hematopoietic stem cells.* Proc Natl Acad Sci U S A, 2010. **107**(17): p. 7799-804.
18. Zhang, J., et al., *Identification of the haematopoietic stem cell niche and control of the niche size.* Nature, 2003. **425**(6960): p. 836-41.
19. Sugiyama, T., et al., *Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches.* Immunity, 2006. **25**(6): p. 977-88.
20. Isern, J. and S. Mendez-Ferrer, *Stem cell interactions in a bone marrow niche.* Curr Osteoporos Rep, 2011. **9**(4): p. 210-8.
21. Sauvageau, G., N.N. Iscove, and R.K. Humphries, *In vitro and in vivo expansion of hematopoietic stem cells.* Oncogene, 2004. **23**(43): p. 7223-32.
22. Coskun, S. and K.K. Hirschi, *Establishment and regulation of the HSC niche: Roles of osteoblastic and vascular compartments.* Birth Defects Res C Embryo Today, 2010. **90**(4): p. 229-42.
23. An, X. and N. Mohandas, *Erythroblastic islands, terminal erythroid differentiation and reticulocyte maturation.* Int J Hematol, 2011. **93**(2): p. 139-43.
24. Chasis, J.A. and N. Mohandas, *Erythroblastic islands: niches for erythropoiesis.* Blood, 2008. **112**(3): p. 470-8.
25. Mohandas, N. and M. Prenant, *Three-dimensional model of bone marrow.* Blood, 1978. **51**(4): p. 633-43.

26. Lee, S.H., et al., *Isolation and immunocytochemical characterization of human bone marrow stromal macrophages in hemopoietic clusters.* J Exp Med, 1988. **168**(3): p. 1193-8.
27. Lee, G., et al., *Targeted gene deletion demonstrates that the cell adhesion molecule ICAM-4 is critical for erythroblastic island formation.* Blood, 2006. **108**(6): p. 2064-71.
28. Friend, C., et al., *Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: stimulation of erythroid differentiation by dimethyl sulfoxide.* Proc Natl Acad Sci U S A, 1971. **68**(2): p. 378-82.
29. Dolznig, H., et al., *Establishment of normal, terminally differentiating mouse erythroid progenitors: molecular characterization by cDNA arrays.* FASEB J, 2001. **15**(8): p. 1442-4.
30. Lozzio, C.B. and B.B. Lozzio, *Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome.* Blood, 1975. **45**(3): p. 321-34.
31. Weiss, M.J., C. Yu, and S.H. Orkin, *Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line.* Mol Cell Biol, 1997. **17**(3): p. 1642-51.
32. Tsang, A.P., et al., *FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation.* Cell, 1997. **90**(1): p. 109-19.
33. Keller, G.M., *In vitro differentiation of embryonic stem cells.* Curr Opin Cell Biol, 1995. **7**(6): p. 862-9.
34. Shen, J. and C.K. Qu, *In vitro hematopoietic differentiation of murine embryonic stem cells.* Methods Mol Biol, 2008. **430**: p. 103-18.
35. Hanspal, M. and J.S. Hanspal, *The association of erythroblasts with macrophages promotes erythroid proliferation and maturation: a 30-kD heparin-binding protein is involved in this contact.* Blood, 1994. **84**(10): p. 3494-504.
36. Hanspal, M., Y. Smockova, and Q. Uong, *Molecular identification and functional characterization of a novel protein that mediates the attachment of erythroblasts to macrophages.* Blood, 1998. **92**(8): p. 2940-50.
37. Sadahira, Y., T. Yoshino, and Y. Monobe, *Very late activation antigen 4-vascular cell adhesion molecule 1 interaction is involved in the formation of erythroblastic islands.* J Exp Med, 1995. **181**(1): p. 411-5.
38. Leimberg, M.J., et al., *Macrophages function as a ferritin iron source for cultured human erythroid precursors.* J Cell Biochem, 2008. **103**(4): p. 1211-8.
39. Chow, A., et al., *CD169(+) macrophages provide a niche promoting erythropoiesis under homeostasis and stress.* Nat Med, 2013. **19**(4): p. 429-36.
40. Ramos, P., et al., *Macrophages support pathological erythropoiesis in polycythemia vera and beta-thalassemia.* Nat Med, 2013. **19**(4): p. 437-45.
41. Clarke, A.R., et al., *Requirement for a functional Rb-1 gene in murine development.* Nature, 1992. **359**(6393): p. 328-30.
42. Lee, E.Y., et al., *Mice deficient for Rb are nonviable and show defects in neurogenesis and haematopoiesis.* Nature, 1992. **359**(6393): p. 288-94.
43. Luo, H., et al., *Disruption of palladin results in neural tube closure defects in mice.* Mol Cell Neurosci, 2005. **29**(4): p. 507-15.
44. Liu, X.S., et al., *Disruption of palladin leads to defects in definitive erythropoiesis by interfering with erythroblastic island formation in mouse fetal liver.* Blood, 2007. **110**(3): p. 870-6.
45. Koury, M.J. and M.C. Bondurant, *Erythropoietin retards DNA breakdown and prevents programmed death in erythroid progenitor cells.* Science, 1990. **248**(4953): p. 378-81.
46. Socolovsky, M., et al., *Negative autoregulation by FAS mediates robust fetal erythropoiesis.* PLoS Biol, 2007. **5**(10): p. e252.
47. Dai, C., et al., *Reduction of cell cycle progression in human erythroid progenitor cells treated with tumour necrosis factor alpha occurs with reduced CDK6 and is partially reversed by CDK6 transduction.* Br J Haematol, 2003. **121**(6): p. 919-27.
48. De Maria, R., et al., *Negative regulation of erythropoiesis by caspase-mediated cleavage of GATA-1.* Nature, 1999. **401**(6752): p. 489-93.
49. Zermati, Y., et al., *Transforming growth factor inhibits erythropoiesis by blocking proliferation and accelerating differentiation of erythroid progenitors.* Exp Hematol, 2000. **28**(8): p. 885-94.

50. von Lindern, M., et al., *The glucocorticoid receptor cooperates with the erythropoietin receptor and c-Kit to enhance and sustain proliferation of erythroid progenitors in vitro.* Blood, 1999. **94**(2): p. 550-9.
51. Broudy, V.C., et al., *Interaction of stem cell factor and its receptor c-kit mediates lodgment and acute expansion of hematopoietic cells in the murine spleen.* Blood, 1996. **88**(1): p. 75-81.
52. Lenox, L.E., J.M. Perry, and R.F. Paulson, *BMP4 and Madh5 regulate the erythroid response to acute anemia.* Blood, 2005. **105**(7): p. 2741-8.
53. Millot, S., et al., *Erythropoietin stimulates spleen BMP4-dependent stress erythropoiesis and partially corrects anemia in a mouse model of generalized inflammation.* Blood, 2010. **116**(26): p. 6072-81.
54. Socolovsky, M., *Molecular insights into stress erythropoiesis.* Curr Opin Hematol, 2007. **14**(3): p. 215-24.
55. Fuda, N.J., M.B. Ardehali, and J.T. Lis, *Defining mechanisms that regulate RNA polymerase II transcription in vivo.* Nature, 2009. **461**(7261): p. 186-92.
56. Core, L.J., J.J. Waterfall, and J.T. Lis, *Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.* Science, 2008. **322**(5909): p. 1845-8.
57. Min, I.M., et al., *Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells.* Genes Dev, 2011. **25**(7): p. 742-54.
58. Chao, S.H. and D.H. Price, *Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo.* J Biol Chem, 2001. **276**(34): p. 31793-9.
59. Ni, Z., et al., *P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo.* Mol Cell Biol, 2008. **28**(3): p. 1161-70.
60. Marshall, N.F., et al., *Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase.* J Biol Chem, 1996. **271**(43): p. 27176-83.
61. Peng, J., et al., *Identification of multiple cyclin subunits of human P-TEFb.* Genes Dev, 1998. **12**(5): p. 755-62.
62. Wada, T., et al., *DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs.* Genes Dev, 1998. **12**(3): p. 343-56.
63. Wada, T., et al., *Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro.* EMBO J, 1998. **17**(24): p. 7395-403.
64. Yamaguchi, Y., et al., *NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation.* Cell, 1999. **97**(1): p. 41-51.
65. Yamada, T., et al., *P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation.* Mol Cell, 2006. **21**(2): p. 227-37.
66. Corden, J.L., et al., *A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II.* Proc Natl Acad Sci U S A, 1985. **82**(23): p. 7934-8.
67. Komarnitsky, P., E.J. Cho, and S. Buratowski, *Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription.* Genes Dev, 2000. **14**(19): p. 2452-60.
68. Marshall, N.F. and D.H. Price, *Purification of P-TEFb, a transcription factor required for the transition into productive elongation.* J Biol Chem, 1995. **270**(21): p. 12335-8.
69. Glover-Cutter, K., et al., *TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II.* Mol Cell Biol, 2009. **29**(20): p. 5455-64.
70. Rasmussen, E.B. and J.T. Lis, *In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 7923-7.
71. McCracken, S., et al., *5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II.* Genes Dev, 1997. **11**(24): p. 3306-18.
72. Kim, E., et al., *Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA.* J Cell Biol, 1997. **136**(1): p. 19-28.
73. Misteli, T. and D.L. Spector, *RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo.* Mol Cell, 1999. **3**(6): p. 697-705.
74. Peterlin, B.M. and D.H. Price, *Controlling the elongation phase of transcription with P-TEFb.*

Mol Cell, 2006. **23**(3): p. 297-305.
75. Ghamari, A., et al., *In vivo live imaging of RNA polymerase II transcription factories in primary cells.* Genes Dev, 2013. **27**(7): p. 767-77.
76. Shi, X., et al., *Paf1p, an RNA polymerase II-associated factor in Saccharomyces cerevisiae, may have both positive and negative roles in transcription.* Mol Cell Biol, 1996. **16**(2): p. 669-76.
77. Shi, X., et al., *Cdc73p and Paf1p are found in a novel RNA polymerase II-containing complex distinct from the Srbp-containing holoenzyme.* Mol Cell Biol, 1997. **17**(3): p. 1160-9.
78. Yart, A., et al., *The HRPT2 tumor suppressor gene product parafibromin associates with human PAF1 and RNA polymerase II.* Mol Cell Biol, 2005. **25**(12): p. 5052-60.
79. Mueller, C.L. and J.A. Jaehning, *Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex.* Mol Cell Biol, 2002. **22**(7): p. 1971-80.
80. Rozenblatt-Rosen, O., et al., *The parafibromin tumor suppressor protein is part of a human Paf1 complex.* Mol Cell Biol, 2005. **25**(2): p. 612-20.
81. Zhu, B., et al., *The human PAF complex coordinates transcription with events downstream of RNA synthesis.* Genes Dev, 2005. **19**(14): p. 1668-73.
82. Mueller, C.L., et al., *The Paf1 complex has functions independent of actively transcribing RNA polymerase II.* Mol Cell, 2004. **14**(4): p. 447-56.
83. Akanuma, T., et al., *Paf1 complex homologues are required for Notch-regulated transcription during somite segmentation.* EMBO Rep, 2007. **8**(9): p. 858-63.
84. Nguyen, C.T., et al., *The PAF1 complex component Leo1 is essential for cardiac and neural crest development in zebrafish.* Dev Biol, 2010. **341**(1): p. 167-75.
85. Langenbacher, A.D., et al., *The PAF1 complex differentially regulates cardiomyocyte specification.* Dev Biol, 2011. **353**(1): p. 19-28.
86. Wang, P., et al., *Parafibromin, a component of the human PAF complex, regulates growth factors and is required for embryonic development and survival in adult mice.* Mol Cell Biol, 2008. **28**(9): p. 2930-40.
87. Adelman, K., et al., *Drosophila Paf1 modulates chromatin structure at actively transcribed genes.* Mol Cell Biol, 2006. **26**(1): p. 250-60.
88. Amrich, C.G., et al., *Cdc73 subunit of Paf1 complex contains C-terminal Ras-like domain that promotes association of Paf1 complex with chromatin.* J Biol Chem, 2012. **287**(14): p. 10863-75.
89. Qiu, H., et al., *Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex.* EMBO J, 2012. **31**(16): p. 3494-505.
90. Krogan, N.J., et al., *The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation.* Mol Cell, 2003. **11**(3): p. 721-9.
91. Chu, Y., et al., *Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes.* EMBO J, 2007. **26**(22): p. 4646-56.
92. Ng, H.H., S. Dole, and K. Struhl, *The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B.* J Biol Chem, 2003. **278**(36): p. 33625-8.
93. Xiao, T., et al., *Histone H2B ubiquitylation is associated with elongating RNA polymerase II.* Mol Cell Biol, 2005. **25**(2): p. 637-51.
94. Piro, A.S., et al., *Small region of Rtf1 protein can substitute for complete Paf1 complex in facilitating global histone H2B ubiquitylation in yeast.* Proc Natl Acad Sci U S A, 2012. **109**(27): p. 10837-42.
95. Rozenblatt-Rosen, O., et al., *The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3′ mRNA processing factors.* Proc Natl Acad Sci U S A, 2009. **106**(3): p. 755-60.
96. Nagaike, T., et al., *Transcriptional activators enhance polyadenylation of mRNA precursors.* Mol Cell, 2011. **41**(4): p. 409-18.
97. Lin, L., et al., *The parafibromin tumor suppressor protein inhibits cell proliferation by repression of the c-myc proto-oncogene.* Proc Natl Acad Sci U S A, 2008. **105**(45): p. 17420-5.
98. Yang, Y.J., et al., *The tumor suppressor, parafibromin, mediates histone H3 K9 methylation for cyclin D1 repression.* Nucleic Acids Res, 2010. **38**(2): p. 382-90.
99. Ding, L., et al., *A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1*

complex for embryonic stem cell identity. Cell Stem Cell, 2009. **4**(5): p. 403-15.
100. Muntean, A.G., et al., *The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis.* Cancer Cell, 2010. **17**(6): p. 609-21.
101. Mosimann, C., G. Hausmann, and K. Basler, *The role of Parafibromin/Hyrax as a nuclear Gli/Ci-interacting protein in Hedgehog target gene control.* Mech Dev, 2009. **126**(5-6): p. 394-405.
102. Tenney, K., et al., *Drosophila Rtf1 functions in histone methylation, gene expression, and Notch signaling.* Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11970-4.
103. Mosimann, C., G. Hausmann, and K. Basler, *Parafibromin/Hyrax activates Wnt/Wg target gene transcription by direct association with beta-catenin/Armadillo.* Cell, 2006. **125**(2): p. 327-41.
104. Takahashi, A., et al., *SHP2 tyrosine phosphatase converts parafibromin/Cdc73 from a tumor suppressor to an oncogenic driver.* Mol Cell, 2011. **43**(1): p. 45-56.
105. Smith, E., C. Lin, and A. Shilatifard, *The super elongation complex (SEC) and MLL in development and disease.* Genes Dev, 2011. **25**(7): p. 661-72.
106. Shattuck, T.M., et al., *Somatic and germ-line mutations of the HRPT2 gene in sporadic parathyroid carcinoma.* N Engl J Med, 2003. **349**(18): p. 1722-9.
107. Newey, P.J., M.R. Bowl, and R.V. Thakker, *Parafibromin--functional insights.* J Intern Med, 2009. **266**(1): p. 84-98.
108. Tolhuis, B., et al., *Looping and interaction between hypersensitive sites in the active beta-globin locus.* Mol Cell, 2002. **10**(6): p. 1453-65.
109. Palstra, R.J., et al., *The beta-globin nuclear compartment in development and erythroid differentiation.* Nat Genet, 2003. **35**(2): p. 190-4.
110. Malik, S. and R.G. Roeder, *The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation.* Nat Rev Genet, 2010. **11**(11): p. 761-72.
111. Koch, F., et al., *Genome-wide RNA polymerase II: not genes only!* Trends Biochem Sci, 2008. **33**(6): p. 265-73.
112. Palis, J., et al., *Primitive erythropoiesis in the mammalian embryo.* Int J Dev Biol, 2010. **54**(6-7): p. 1011-8.
113. Silver, L. and J. Palis, *Initiation of murine embryonic erythropoiesis: a spatial analysis.* Blood, 1997. **89**(4): p. 1154-64.
114. Martin, D.I., et al., *Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages.* Nature, 1990. **344**(6265): p. 444-7.
115. Ito, E., et al., *Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis.* Nature, 1993. **362**(6419): p. 466-8.
116. Fujiwara, Y., et al., *Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1.* Proc Natl Acad Sci U S A, 1996. **93**(22): p. 12355-8.
117. Whyatt, D., et al., *An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells.* Nature, 2000. **406**(6795): p. 519-24.
118. Takahashi, S., et al., *Arrest in primitive erythroid cell development caused by promoter-specific disruption of the GATA-1 gene.* J Biol Chem, 1997. **272**(19): p. 12611-5.
119. Mucenski, M.L., et al., *A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis.* Cell, 1991. **65**(4): p. 677-89.
120. Ramsay, R.G. and T.J. Gonda, *MYB function in normal and cancer cells.* Nat Rev Cancer, 2008. **8**(7): p. 523-34.
121. Weinstein, B.M., et al., *Hematopoietic mutations in the zebrafish.* Development, 1996. **123**: p. 303-9.
122. Kobayashi-Osaki, M., et al., *GATA motifs regulate early hematopoietic lineage-specific expression of the Gata2 gene.* Mol Cell Biol, 2005. **25**(16): p. 7005-20.
123. Pai, S.Y., et al., *Critical roles for transcription factor GATA-3 in thymocyte development.* Immunity, 2003. **19**(6): p. 863-75.
124. Kouros-Mehr, H., et al., *GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland.* Cell, 2006. **127**(5): p. 1041-55.
125. Nardelli, J., et al., *Expression and genetic interaction of transcription factors GATA-2 and GATA-3 during development of the mouse central nervous system.* Dev Biol, 1999. **210**(2): p.

305-21.
126. Kornhauser, J.M., et al., *Temporal and spatial changes in GATA transcription factor expression are coincident with development of the chicken optic tectum.* Brain Res Mol Brain Res, 1994. **23**(1-2): p. 100-10.
127. Pevny, L., et al., *Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1.* Nature, 1991. **349**(6306): p. 257-60.
128. Iwasaki, H., et al., *GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages.* Immunity, 2003. **19**(3): p. 451-62.
129. Rodriguez, P., et al., *GATA-1 forms distinct activating and repressive complexes in erythroid cells.* EMBO J, 2005. **24**(13): p. 2354-66.
130. Chlon, T.M., L.C. Dore, and J.D. Crispino, *Cofactor-mediated restriction of GATA-1 chromatin occupancy coordinates lineage-specific gene expression.* Mol Cell, 2012. **47**(4): p. 608-21.
131. Hasegawa, A., et al., *Mature erythrocyte membrane homeostasis is compromised by loss of the GATA1-FOG1 interaction.* Blood, 2012. **119**(11): p. 2615-23.
132. Anguita, E., et al., *Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2.* EMBO J, 2004. **23**(14): p. 2841-52.
133. Lahlil, R., et al., *SCL assembles a multifactorial complex that determines glycophorin A expression.* Mol Cell Biol, 2004. **24**(4): p. 1439-52.
134. Soler, E., et al., *The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation.* Genes Dev, 2010. **24**(3): p. 277-89.
135. Yu, M., et al., *Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis.* Mol Cell, 2009. **36**(4): p. 682-95.
136. Pope, N.J. and E.H. Bresnick, *Establishment of a cell-type-specific genetic network by the mediator complex component Med1.* Mol Cell Biol, 2013. **33**(10): p. 1938-55.
137. Grass, J.A., et al., *GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8811-6.
138. Kaneko, H., R. Shimizu, and M. Yamamoto, *GATA factor switching during erythroid differentiation.* Curr Opin Hematol, 2010. **17**(3): p. 163-8.
139. Vicente, C., et al., *The role of the GATA2 transcription factor in normal and malignant hematopoiesis.* Crit Rev Oncol Hematol, 2012. **82**(1): p. 1-17.
140. Ferreira, R., et al., *Dynamic regulation of Gata factor levels is more important than their identity.* Blood, 2007. **109**(12): p. 5481-90.
141. Song, S.H., C. Hou, and A. Dean, *A positive role for NLI/Ldb1 in long-range beta-globin locus control region function.* Mol Cell, 2007. **28**(5): p. 810-22.
142. Deng, W., et al., *Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor.* Cell, 2012. **149**(6): p. 1233-44.
143. Cross, A.J., et al., *LIM domain binding proteins 1 and 2 have different oligomeric states.* J Mol Biol, 2010. **399**(1): p. 133-44.
144. Matthews, J.M. and J.E. Visvader, *LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins.* EMBO Rep, 2003. **4**(12): p. 1132-7.
145. Mukhopadhyay, M., et al., *Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation.* Development, 2003. **130**(3): p. 495-505.
146. Meier, N., et al., *Novel binding partners of Ldb1 are required for haematopoietic development.* Development, 2006. **133**(24): p. 4913-23.
147. Brand, M., et al., *Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics.* Nat Struct Mol Biol, 2004. **11**(1): p. 73-80.
148. Li, L., et al., *Nuclear adaptor Ldb1 regulates a transcriptional program essential for the maintenance of hematopoietic stem cells.* Nat Immunol, 2011. **12**(2): p. 129-36.
149. Dey-Guha, I., et al., *Role of ldb1 in adult intestinal homeostasis.* Int J Biol Sci, 2009. **5**(7): p. 686-94.
150. Storbeck, C.J., et al., *The Ldb1 and Ldb2 transcriptional cofactors interact with the Ste20-like kinase SLK and regulate cell migration.* Mol Biol Cell, 2009. **20**(19): p. 4174-82.
151. Jan, Y.N. and L.Y. Jan, *HLH proteins, fly neurogenesis, and vertebrate myogenesis.* Cell, 1993. **75**(5): p. 827-30.

152. Begley, C.G., et al., *The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif.* Proc Natl Acad Sci U S A, 1989. **86**(24): p. 10128-32.
153. Robb, L., et al., *Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene.* Proc Natl Acad Sci U S A, 1995. **92**(15): p. 7075-9.
154. Ismailoglu, I., et al., *Mesodermal patterning activity of SCL.* Exp Hematol, 2008. **36**(12): p. 1593-603.
155. Mead, P.E., et al., *SCL specifies hematopoietic mesoderm in Xenopus embryos.* Development, 1998. **125**(14): p. 2611-20.
156. Warren, A.J., et al., *The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development.* Cell, 1994. **78**(1): p. 45-57.
157. Patterson, L.J., et al., *The transcription factors Scl and Lmo2 act together during development of the hemangioblast in zebrafish.* Blood, 2007. **109**(6): p. 2389-98.
158. Hahm, K., et al., *Defective neural tube closure and anteroposterior patterning in mice lacking the LIM protein LMO4 or its interacting partner Deaf-1.* Mol Cell Biol, 2004. **24**(5): p. 2074-82.
159. Lee, S.K., et al., *The LIM domain-only protein LMO4 is required for neural tube closure.* Mol Cell Neurosci, 2005. **28**(2): p. 205-14.
160. Okuda, T., et al., *AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis.* Cell, 1996. **84**(2): p. 321-30.
161. North, T., et al., *Cbfa2 is required for the formation of intra-aortic hematopoietic clusters.* Development, 1999. **126**(11): p. 2563-75.
162. Ichikawa, M., et al., *AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis.* Nat Med, 2004. **10**(3): p. 299-304.
163. van Riel, B., et al., *A novel complex, RUNX1-MYEF2, represses hematopoietic genes in erythroid cells.* Mol Cell Biol, 2012. **32**(19): p. 3814-22.
164. Richmond, T.D., M. Chohan, and D.L. Barber, *Turning cells red: signal transduction mediated by erythropoietin.* Trends Cell Biol, 2005. **15**(3): p. 146-55.
165. D'Andrea, A.D., H.F. Lodish, and G.G. Wong, *Expression cloning of the murine erythropoietin receptor.* Cell, 1989. **57**(2): p. 277-85.
166. Wu, H., et al., *Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor.* Cell, 1995. **83**(1): p. 59-67.
167. Ingley, E., et al., *Lyn deficiency reduces GATA-1, EKLF and STAT5, and induces extramedullary stress erythropoiesis.* Oncogene, 2005. **24**(3): p. 336-43.
168. Schmidt, U., et al., *Btk is required for an efficient response to erythropoietin and for SCF-controlled protection against TRAIL in erythroid progenitors.* J Exp Med, 2004. **199**(6): p. 785-95.
169. Wojchowski, D.M., et al., *Signal transduction in the erythropoietin receptor system.* Exp Cell Res, 1999. **253**(1): p. 143-56.
170. Starr, R., et al., *A family of cytokine-inducible inhibitors of signalling.* Nature, 1997. **387**(6636): p. 917-21.
171. Neel, B.G., H. Gu, and L. Pao, *The 'Shp'ing news: SH2 domain-containing tyrosine phosphatases in cell signaling.* Trends Biochem Sci, 2003. **28**(6): p. 284-93.
172. Tauchi, T., et al., *Tyrosine 425 within the activated erythropoietin receptor binds Syp, reduces the erythropoietin required for Syp tyrosine phosphorylation, and promotes mitogenesis.* Blood, 1996. **87**(11): p. 4495-501.
173. Irie-Sasaki, J., et al., *CD45 is a JAK phosphatase and negatively regulates cytokine receptor signalling.* Nature, 2001. **409**(6818): p. 349-54.
174. Myers, M.P., et al., *TYK2 and JAK2 are substrates of protein-tyrosine phosphatase 1B.* J Biol Chem, 2001. **276**(51): p. 47771-4.
175. Scholl, C., D.G. Gilliland, and S. Frohling, *Deregulation of signaling pathways in acute myeloid leukemia.* Semin Oncol, 2008. **35**(4): p. 336-45.
176. Buehr, M., et al., *Proliferation and migration of primordial germ cells in We/We mouse embryos.* Dev Dyn, 1993. **198**(3): p. 182-9.
177. Mintz, B. and E.S. Russell, *Gene-induced embryological modifications of primordial germ cells*

in the mouse. J Exp Zool, 1957. **134**(2): p. 207-37.
178. Russel, E.S., M.W. Thompson, and E. McFarland, *Analysis of effects of W and f genic substitutions on fetal mouse hematology.* Genetics, 1968. **58**(2): p. 259-70.
179. Russell, E.S. and E.L. Fondal, *Quantitative analysis of the normal and four alternative degrees of an inherited macrocytic anemia in the house mouse. I. Number and size of erythrocytes.* Blood, 1951. **6**(10): p. 892-905.
180. Bernex, F., et al., *Spatial and temporal patterns of c-kit-expressing cells in WlacZ/+ and WlacZ/WlacZ mouse embryos.* Development, 1996. **122**(10): p. 3023-33.
181. Kent, D., et al., *Regulation of hematopoietic stem cells by the steel factor/KIT signaling pathway.* Clin Cancer Res, 2008. **14**(7): p. 1926-30.
182. Gutierrez, L., et al., *Homotypic signalling regulates Gata1 activity in the erythroblastic island.* Development, 2004. **131**(13): p. 3183-93.
183. Amatruda, J.F. and L.I. Zon, *Dissecting hematopoiesis and disease using the zebrafish.* Dev Biol, 1999. **216**(1): p. 1-15.
184. Davidson, A.J. and L.I. Zon, *The 'definitive' (and 'primitive') guide to zebrafish hematopoiesis.* Oncogene, 2004. **23**(43): p. 7233-46.
185. Paik, E.J. and L.I. Zon, *Hematopoietic development in the zebrafish.* Int J Dev Biol, 2010. **54**(6-7): p. 1127-37.
186. Kimmel, C.B., R.M. Warga, and T.F. Schilling, *Origin and organization of the zebrafish fate map.* Development, 1990. **108**(4): p. 581-94.
187. Lieschke, G.J., et al., *Zebrafish SPI-1 (PU.1) marks a site of myeloid development independent of primitive erythropoiesis: implications for axial patterning.* Dev Biol, 2002. **246**(2): p. 274-95.
188. Bertrand, J.Y., et al., *Definitive hematopoiesis initiates through a committed erythromyeloid progenitor in the zebrafish embryo.* Development, 2007. **134**(23): p. 4147-56.
189. Thompson, M.A., et al., *The cloche and spadetail genes differentially affect hematopoiesis and vasculogenesis.* Dev Biol, 1998. **197**(2): p. 248-69.
190. Murayama, E., et al., *Tracing hematopoietic precursor migration to successive hematopoietic organs during zebrafish development.* Immunity, 2006. **25**(6): p. 963-75.
191. Patterson, L.J., M. Gering, and R. Patient, *Scl is required for dorsal aorta as well as blood formation in zebrafish embryos.* Blood, 2005. **105**(9): p. 3502-11.
192. Galloway, J.L., et al., *Loss of gata1 but not gata2 converts erythropoiesis to myelopoiesis in zebrafish embryos.* Dev Cell, 2005. **8**(1): p. 109-16.
193. Tsai, F.Y., et al., *An early haematopoietic defect in mice lacking the transcription factor GATA-2.* Nature, 1994. **371**(6494): p. 221-6.
194. Stainier, D.Y., et al., *Cloche, an early acting zebrafish gene, is required by both the endothelial and hematopoietic lineages.* Development, 1995. **121**(10): p. 3141-50.
195. Kelsh, R.N., et al., *Zebrafish pigmentation mutations and the processes of neural crest development.* Development, 1996. **123**: p. 369-89.
196. Ransom, D.G., et al., *The zebrafish moonshine gene encodes transcriptional intermediary factor 1gamma, an essential regulator of hematopoiesis.* PLoS Biol, 2004. **2**(8): p. E237.
197. Kim, J. and V. Kaartinen, *Generation of mice with a conditional allele for Trim33.* Genesis, 2008. **46**(6): p. 329-33.
198. Bai, X., et al., *TIF1gamma controls erythroid cell fate by regulating transcription elongation.* Cell, 2010. **142**(1): p. 133-43.
199. Fields, S. and O. Song, *A novel genetic system to detect protein-protein interactions.* Nature, 1989. **340**(6230): p. 245-6.
200. Topper, J.N., et al., *CREB binding protein is a required coactivator for Smad-dependent, transforming growth factor beta transcriptional responses in endothelial cells.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9506-11.
201. Estojak, J., R. Brent, and E.A. Golemis, *Correlation of two-hybrid affinity data with in vitro measurements.* Mol Cell Biol, 1995. **15**(10): p. 5820-9.
202. Lievens, S., I. Lemmens, and J. Tavernier, *Mammalian two-hybrids come of age.* Trends Biochem Sci, 2009. **34**(11): p. 579-88.
203. Barrios-Rodiles, M., et al., *High-throughput mapping of a dynamic signaling network in mammalian cells.* Science, 2005. **307**(5715): p. 1621-5.

204. Caufield, J.H., N. Sakhawalkar, and P. Uetz, *A comparison and optimization of yeast two-hybrid systems.* Methods, 2012. **58**(4): p. 317-24.
205. Förster, T., *Intermolecular energy migration and fluorescence.* Ann. Phys., 1948. **437**(1-2): p. 55-75.
206. Xu, Y., D.W. Piston, and C.H. Johnson, *A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins.* Proc Natl Acad Sci U S A, 1999. **96**(1): p. 151-6.
207. Padilla-Parra, S. and M. Tramier, *FRET microscopy in the living cell: different approaches, strengths and weaknesses.* Bioessays, 2012. **34**(5): p. 369-76.
208. Hu, C.D., Y. Chinenov, and T.K. Kerppola, *Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation.* Mol Cell, 2002. **9**(4): p. 789-98.
209. Kodama, Y. and C.D. Hu, *Bimolecular fluorescence complementation (BiFC): a 5-year update and future perspectives.* Biotechniques, 2012. **53**(5): p. 285-98.
210. Fredriksson, S., et al., *Protein detection using proximity-dependent DNA ligation assays.* Nat Biotechnol, 2002. **20**(5): p. 473-7.
211. Gustafsdottir, S.M., et al., *Proximity ligation assays for sensitive and specific protein analyses.* Anal Biochem, 2005. **345**(1): p. 2-9.
212. Soderberg, O., et al., *Direct observation of individual endogenous protein complexes in situ by proximity ligation.* Nat Methods, 2006. **3**(12): p. 995-1000.

Chapter 2

The temporal and spatial emergence of GATA1 complexes in developing mouse embryos

The temporal and spatial emergence of GATA1 complexes in developing mouse embryos

Xiao Yu, Andrea Martella, Mary Stevens, Ralph Stadhouders, Frank G. Grosveld and Charlotte Andrieu-Soler.

Abstract

Hematopoietic stem cells (HSCs) give rise to all hematopoietic lineages. Intra-embryonic hematopoietic cells appear in the aorta-gonad-mesonephros (AGM) region of mouse embryos at embryonic day 9 (E9), migrate to the fetal liver at E10 and finally colonize the bone marrow after E15 where HSCs stay present throughout adult life. The transcription factor gata1 is an essential regulator of hematopoiesis in mouse, and also in zebrafish. During erythroid differentiation GATA1 forms different complexes with other transcription factors such as LDB1, TAL1, E2A and LMO2 (known as the LDB1 complex) or with FOG1. The functions of the GATA1 complexes have been extensively studied in definitive erythroid differentiation; however, the temporal and spatial formation of these complexes during embryonic hematopoiesis is still unknown. We therefore applied PLA (proximity ligation assay) to detect, localize and quantify individual interactions in mouse embryonic tissue slices. We identified a CD71+/TER119- cell population in fetal liver that first shows the interaction between GATA1 and LDB1 as a part of a protein complex able to activate the remaining erythroid differentiation program. The results show that the activation of the LDB1 complex occurs quite late, i.e. at the proerythroblast stage of erythroid differentiation, and confirm the power of PLA in studying the dynamic interaction of proteins in cell differentiation.

Introduction

In mouse embryonic development, the first hematopoietic cells appear in blood islands in the yolk sac on embryonic day 6.5 (E6.5). Later, and only at E10.5 to E11, hematopoietic stem cells (HSCs), defined here as cells that confer complete, long-term, multilineage, substantial repopulation of irradiated adult recipient mice, appear in the aorta-gonad-mesonephros (AGM) region within the embryo (and in the vitelline and umbilical arteries). These cells migrate to the fetal liver, mature from pre-HSC to HSC, and eventually reside in the adult bone marrow [1], although different models for the origin of definitive HSCs have been proposed [2]. One of the lineages originating from HSCs is the erythroid lineage that establishes and maintains the red cell population.

The DNA-binding transcription factor GATA1 is one of the essential proteins for the erythropoietic and megakaryocytic programs. *Gata1* knockout embryos die between E9.5 to E10 due to a block of primitive and definitive erythroid cell differentiation at the proerythroblast stage, leading to the absence of mature red blood cells. GATA1 can form several complexes with other proteins to regulate erythroid gene expression. Two proteins of particular interest bind directly to GATA1. The first one, FOG1 (Friend of GATA1), binds to the N-terminal zinc finger (ZnF) of GATA1 and recruits the chromatin remodeling complex NuRD/MeCP1 and/or the C-terminal binding protein (CTBP) corepressor-containing complex to regulate GATA1 target genes [3]. The second protein, the ubiquitously produced LIM-domain-binding protein 1 (LDB1) functions as a scaffold protein to form multiprotein transcription complexes to regulate the differentiation of various cell types. Due to interactions of LDB1 with many other transcription factors, *Ldb1* knockout mice die between E9.5 and E10 due to severe defects in a number of developing embryogenic tissues, including abnormal hematopoietic cell development [4]. This abnormal hematopoiesis has also been observed in knockout mouse embryos lacking the LDB1 binding-partners TAL1 or LMO2.

Despite the emerging knowledge on GATA1 binding partners, it is not known when and where GATA1 complexes are formed. In order to identify the temporal and spatial appearance of GATA1/FOG1 complex and GATA1/LDB1 complex, we applied the proximity ligation assay (PLA), originally developed by Ulf and co-workers [5], in differentiated mouse ES cells and fetal liver cells. We detect significant GATA1/LDB1 interaction in CD71+ cells of the fetal liver. Knockdown of LDB1 *in vitro* led to fetal cell death and decreased the CD71+ cell populations, providing functional evidence for its essential role at that stage of erythroid differentiation in normal fetal liver.

Results

LDB1 complexes during *in vitro* ES cell differentiation

Previous LDB1 immunoprecipitation (IP) experiments in ES cells showed that LDB1 does not bind to GATA1, even though steady-state mRNAs encoding GATA1 are expressed at very low level (unpublished data). First, we wanted to test when the LDB1/GATA1 complex forms during *in vitro* mouse ES cell differentiation. *In vitro* differentiation of ES cells leads to the synchronous temporal and spatial emergence of endothelial cells and blood cells, via the common precursor for these two lineages, the hemangioblast or blast colony-forming cells (BL-CFCs) [6-8]. BL-CFCs appear at day 4 of ES cell differentiation; however, the generation of HSCs from ES cell differentiation has not been successful. We therefore tested whether the two GATA1 complexes are already formed at this stage. We performed an IP for LDB1 in nuclear extracts from day 4 and 5 ES cells, differentiated in embryoid bodies (EBs) (**Fig1-a**). The E2A proteins could be detected in the LDB1 IP, although the intensity of the E2A bands at day 4 is very weak. However, LDB1, GATA1 and FOG1 IPs in cells at these time points did not detect their each other. GATA1 detection in GATA1 IPs is weak and only visible at day 5 (as shown in **Fig1-b**, arrow), not day 4. This is probably due to the overall low amount of the protein (**Fig1-b**). The positive controls in MEL cells show that in the LDB1 and FOG1 IPs GATA1 could be detected, whereas a LDB1 IP does not detect FOG1.

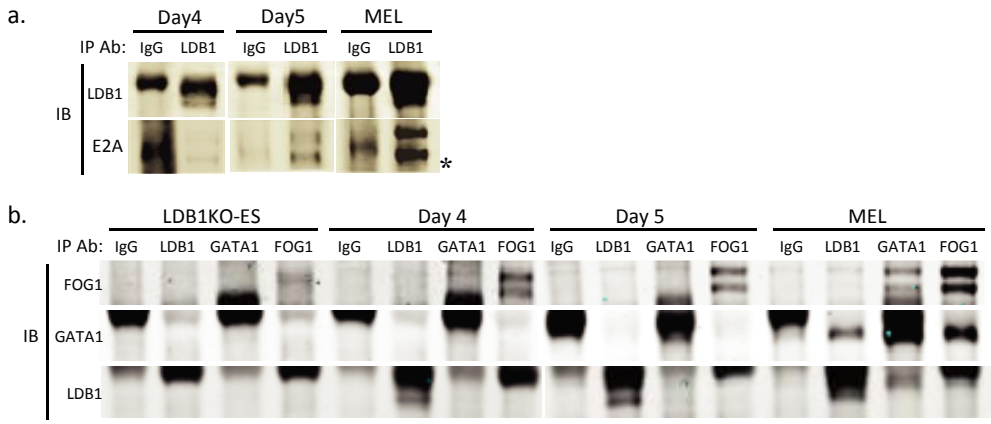


Figure 1. The LDB1 complex during the ES cell differentiated EBs.
a. LDB1 IP on day 4 and day 5 EBs. The IgG control and LDB1 IPs are carried out on nuclear extracts from day 4 and day 5 EBs and MEL cells. Both the LDB1 and E2A bands can be detected. *:background IgG band. **b**. LDB1, GATA1 and FOG1 IPs on day 4 and day 5 EBs, LDB1KO-ES cells as a negative control and MEL cells as a positive control. *:background IgG band. Arrow shows the GATA1 band.

We also applied size-exclusion chromatography to distinguish amongst the different GATA1 complexes. The bulk of GATA1 is not in the same fraction with neither FOG1 nor LDB1 in extracts made from day-4 and day-5 cells, while GATA1 and E2A are present in overlapping fractions (i.e. fractions29 and 30, in D4 and D5 NE) (Supplementary data, **FigS1**). Therefore, we could not confirm the GATA1 complexes at these stages of ES cell differentiation in this assay. Clearly LDB1 already appears to bind to E2A in day-4 cells while it is not clear whether there is GATA1 binding. The positive control MEL cells show that GATA1 overlaps with both LDB1 and FOG1 to form different complexes.

To further identify the binding partners of GATA1, LDB1 and FOG1 in day-4 and 5 cells, we carried out co-IP for these factors and performed mass spectrometry (MS). Due to the low GATA1 levels, GATA1 was not detected by mass spectrometry (even when the level is high trypsin produces only one peptide that can be detected and hence the mascot score will be low). **Table S1** shows the result of the LDB1-IP and FOG1-IP mass spectrometry at day 4 and day 5. The LDB1-IP shows the SSBPs, LMO4 and LDB2 proteins. These three types of protein are also detected in the LDB1-IP in FLK1+ cells (unpublished data). In the LDB1-IP at day 5, the NuRD complex components HDAC2, MTA2/3, CHD4 are detected. HDAC2 can also form another complex together with SIN3A and form the SIN3 HDAC complex. Interestingly the LDB1-IP detects LHX1, a LIM domain protein that is essential for head formation during early development of the mouse [9, 10]. Compared to the LDB1-IP, the FOG1-IP proteins have already formed multi-protein complexes at day 4 including with the NuRD complex, the SIN3 HDAC complex, BHC complex and SWI/SNF complex. Compared to day 4, the MS data indicate that the Mediator complex has partly dissociated from FOG1 on day 5. We could neither detect LDB1 in the FOG1-IP nor the FOG1 complex in the LDB1-IP, confirming that LDB1 and FOG1 are not within the same complex on day 4 or day 5 of ES cell differentiation.

Next, we carried out a Gene Ontology analysis to understand the function of the LDB1 and FOG1 binding partners. This revealed that both the LDB1 and FOG1 partners may be involved in general cellular functions, such as RNA post-transcription modification, gene expression and protein synthesis (**Table S2)**. Compared to this result, some tissue specific development mechanisms are found on day 5, for instance, hematological system development (specifically in the LDB1-IP gene list), skeletal and muscular system development (both in the LDB1-IP and FOG1-IP gene lists) and connective tissue development (specifically in the FOG1-IP gene list).

We conclude that LDB1 and FOG1 already form different complexes early during ES differentiation. They regulate general cellular mechanisms on day 4 and, following further differentiation, they also acquire a more specific role in the development of different cell lineages.

LDB1 complexes during *in vitro* ES cell differentiation

In order to investigate at what time point in ES cell differentiation LDB1/GATA1, GATA1/FOG1 and LDB1/E2A interactions occur, we applied PLA [5] on sliced *in vitro* ES cell differentiated EBs. Two primary antibodies for the respective combinations of transcription factors (see above) were added to the sliced wild-type or *Ldb1* KO EBs, followed by incubation with a secondary antibody (specifically directed against each of the primary antibodies), the latter being conjugated to oligonucleotides (used for amplification). After performing the PLA, the pictures were analyzed using the blobfinder software [11]. This quantifies the number of PLA-positive dots in a cell and the density of the dots related to its nuclear area. Comparing the signals obtained with two primary antibodies versus the negative controls (using only one primary antibody or only secondary antibodies), we plotted the PLA signals as shown in boxplot (**Fig2**). We also checked the potential GATA1/LDB1, GATA1/FOG1 and LDB1/E2A interactions in wild-type ES cell and *Ldb1* KO ES cell differentiated EBs at day 9. The quantification of the PLA signals showed that the GATA1/LDB1 and LDB1/E2A interactions already take place at day 4 of ES cell differentiation. The GATA1/FOG1 interaction appeared 24 hours later. These GATA1/LDB1 and GATA1/FOG1 interactions were not observed in the previously IP and size-exclusion chromatography experiments shown above. This discrepancy is most likely due to the low level of these proteins as well as interactions in the EBs, where it also should be noted that these inevitably contain various cell types. It shows the advantage of PLA in detecting a low amount of proteins and their interaction. The control LDB1 KO cells show no signal.

In conclusion, the PLA assay allowed us to monitor the dynamic changes of different protein complexes even when a very low amount is present during the embryonic development. The PLA signals in EBs are fairly evenly distributed in a subpopulation of the cells, which suggest that there is already a specification between positive and negative cells and that the populations are mixed in these three-dimensional cell aggregates.

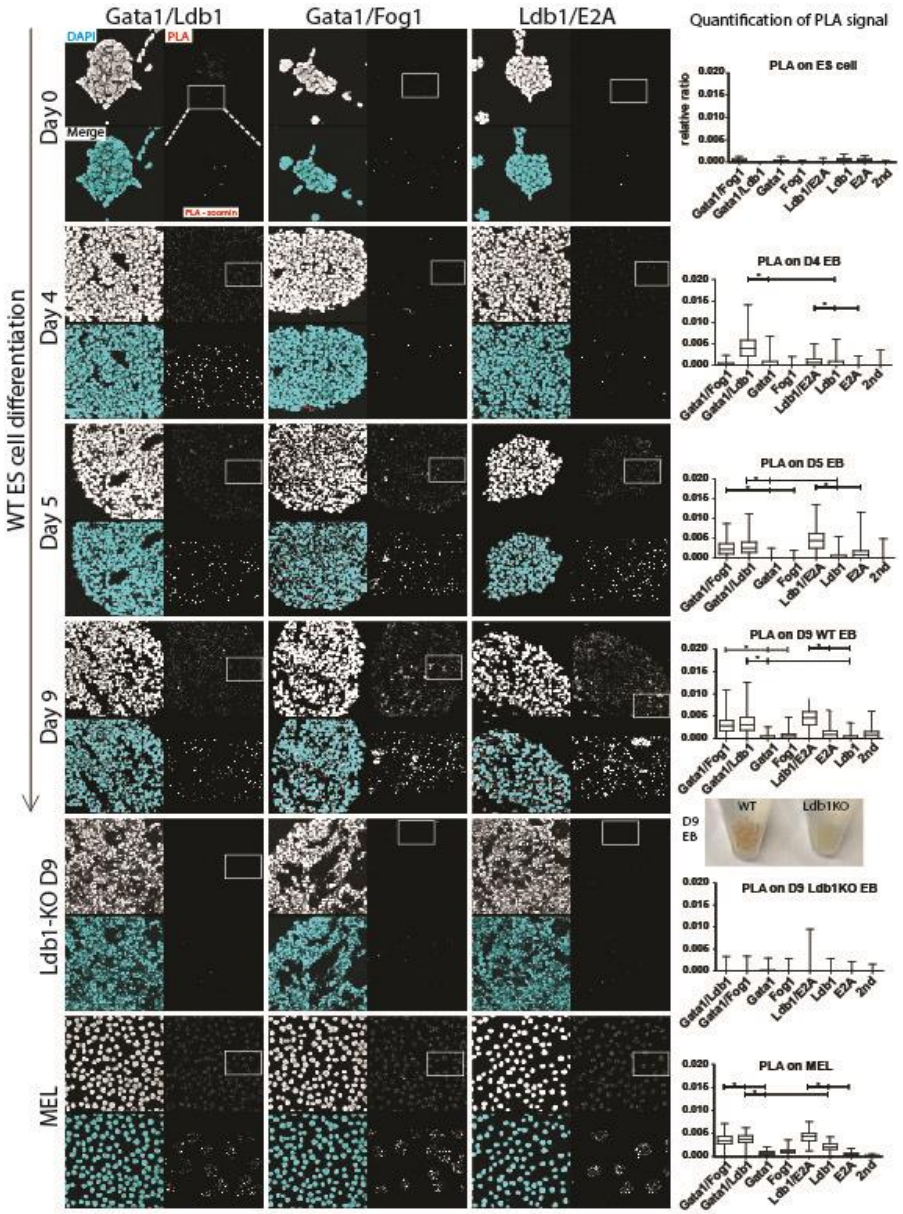


Figure 2. PLA on WT and LDB1-KO EBs on different days

PLA for GATA1/LDB1, GATA1/FOG1 and LDB1/E2A interactions was performed on WT ES cell differentiation at day 0 (ES cells), day 4, day 5 and day9 and LDB1-KO ES cell differentiation at day9. MEL cells as a positive control. The nucleus of the cells was stained with DAPI (top and bottom left of each panel), and PLA signal was in Texas-Red (top right of each panel) . The squared region of the PLA signal has been enlarged to demonstrate the signal (bottom right of each panel). PLA signals have been quantified on each day and compared with negative controls of single primary antibody or secondary antibody alone. * shows the significant interactions between the two primary antibodies PLA signal and single primary antibody controls. The significance was analyzed with Kruskal–Wallis test for the variance.

GATA1 complexes in mouse E12.5 fetal liver

HSCs have moved from the aorta to the fetal liver at E10.5 to E11.5 where they give rise to the different blood cell lineages. In order to understand the temporal appearance of the two GATA1 complexes in the definitive blood cells in the fetal liver, we applied PLA to embryonic tissue sections at E12.5. We compared the GATA1/LDB1 complex (**Fig3-a**, upper panel) and the GATA1/FOG1 complex (lower panel). Each red dot in the image represents a fluorescent signal of GATA1 complex formation. Although the fetal liver tissue is compact and single cells are difficult to distinguish, it is still obvious that in the GATA1/LDB1 PLA some cells contained very dense PLA signals when compared to the other surrounding cells (**Fig3-a**). A similar result is also found in the fetal aorta (data not shown). Specific PLA signals are also detected for another interaction, i.e. LDB1/LMO2, which is part of the same complex of GATA1/LDB1, in fetal liver and also in the dorsal aorta (**Fig3-b**). This suggests that the complete GATA1/LDB1/LMO2 complex is present at high levels in a subpopulation of the cells in fetal liver and aorta. In contrast, the GATA1/FOG1 PLA signal appears more evenly distributed, but is weaker. In agreement with the PLA results, the staining for individual LDB1, GATA1 and FOG1 proteins in the fetal liver tissue sections showed high GATA1 and LDB1 signals in a subpopulation of fetal liver cells (Supplementary data, **FigS2**). The FOG1 staining was very poor to absent, which could be due to the antibody or to the inappropriate 3D arrangement or distance between the GATA1 and FOG1 epitopes. Nevertheless, these results suggest that GATA1/LDB1/LMO2 interaction becomes very important for one or more specific cell types in the fetal liver and aorta. This result would be in agreement with the data showing that the level of GATA1 increases before the end stage of erythroid differentiation [12] and that this increase is primarily found with the LDB1 complex.

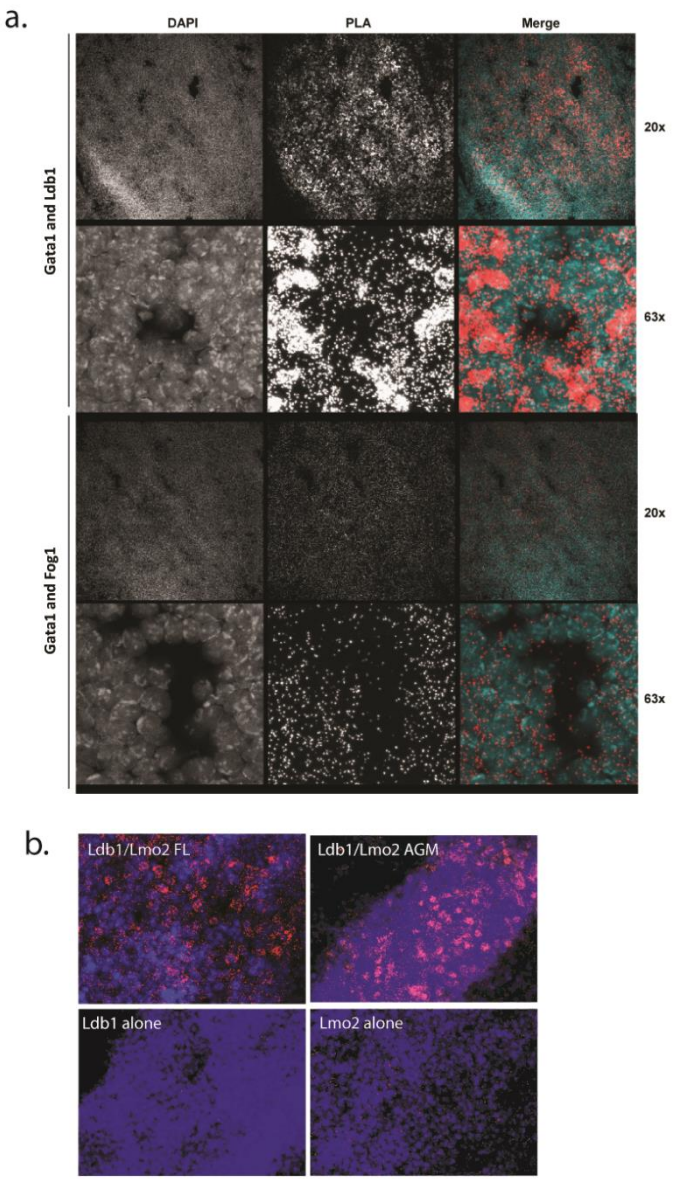


Figure 3. PLA of GATA1 complexes in E12.5 embryonic tissues

a. PLA for Gata1/LDB1 (upper panel) and Gata1/Fog1 (lower panel) interactions was performed on sections of mouse fetal liver tissue at E12.5. DAPI staining was used to visualize the nucleus (cyan), PLA protein interactions were visualized in red. For each interaction, microscopy pictures were made under 20x and 63x optical lens.

b. LDB1/LMO2 interaction in fetal liver (FL) and aorta (AGM) (40x). PLA and DAPI signals are merged in each picture. The Ldb1 alone and Lmo2 alone PLA negative controls were performed on fetal liver tissue using a labeled secondary antibody.

PLA detection of GATA1/LDB1 complex in sorted fetal liver cells

In order to identify which cell populations contain the high PLA signals in the fetal liver, we FACS sorted the cells using two cell surface markers, TER119 and CD71, to distinguish between the different stages of erythropoiesis (from proerythroblasts to orthochromatic erythroblasts) (**Fig4-a**). The cells were separated into three populations, indicated as P2 to P4. P1 contains a population negative for both CD71 and TER119 and represents the early precursor population and all other lineage cell types. **Figure4-b** shows the PLA signal on the P1 to P4 cell populations. The signals are very high in P2 (CD71+/TER119-) and P3 (CD71+/TER119+) populations. Quantification confirmed that the P2 population significantly had the highest PLA signals, while we also see the similar results in the "density of dots per cell", a different measurement due to the different size of the nucleus (data not shown). We also detected the GATA1/LDB1 interaction in PLA experiment increases upon MEL differentiation (**Fig4-d**), which fits with our previous results [12]. These results indicate that MEL cell differentiation is similar to the liver erythroid cell differentiation. MEL non-induced cells represent the P1 population, and induced cells represent P2 (and/or further differentiated) populations. We therefore conclude that the LDB1 and GATA1 proteins interact at the proerythroblast and basophilic erythroblast stages and that this interaction decreases during the final stages of erythroid differentiation *in vivo*. Thus, the function of the GATA1/LDB1 complex appears most important at the middle stages to activate or derepress its target genes in erythroid cells. For instance, in the case of the *Klf1* gene it has been shown (using chromatin immunoprecipitation) that the GATA1/LDB1 complex is binding to its upstream enhancer region and activates *Klf1* in the induced MEL cells [13], while the *α-globin* [14] is also activated by the same complex on its promoter. *Gypa* gene is repressed by the TAL1/ETO2 complex, while during the erythroid differentiation reactivation of this gene occurs [15]. Similar repression of TAL/LSD1 complex in the undifferentiated stage was also found *Epb4.2* [16]. In conclusion, the GATA1/LDB1 complex activates erythroid specific genes at the proerythroblast (P2) stage of erythroid differentiation.

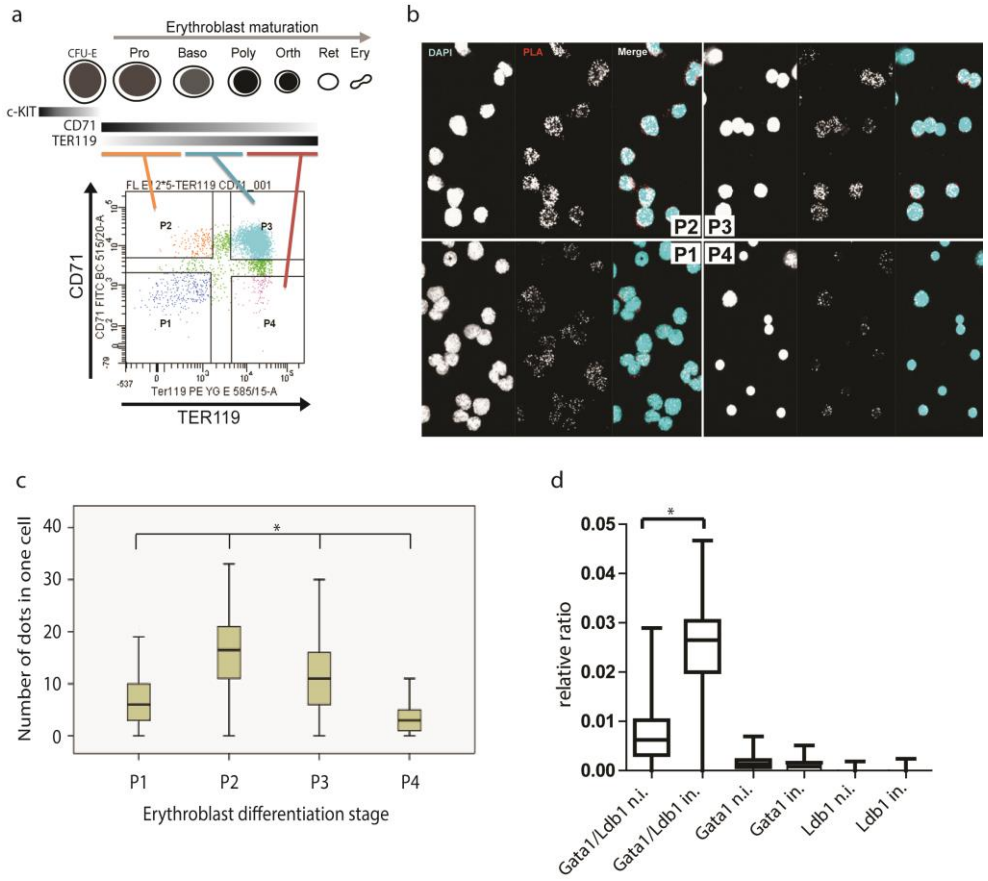


Figure 4. PLA detection of GATA1 complexes on sorted fetal liver cells and MEL cells

a. Schematic description of erythroblast development. Pro, immature proerythroblast; Baso, basophilic erythroblast; Poly, polychromatic erythroblast; Orth, orthochromatic erythroblast; Ret, reticulocytes; Ery, erythrocyte. The cell sorting is based on membrane markers TER119 and CD71. Gray bars represent c-KIT, CD71 and TER119 gene expression.

b. GATA1/LDB1 PLA on the four sorted cell populations. DAPI stains the nucleus in cyan, PLA is in red.

c. Boxplot comparison of PLA signals of "number of dots in one cell" in these cell populations. * Indicates significance which was determined with the Kruskal–Wallis test.

d. Quantification of GATA1/LDB1 PLA signals in MEL non-induced (n.i.) and induced (in.) cells for Gata1/Ldb1, Gata1 antibody alone and Ldb1 antibody alone negative controls. Significance was determined with the Kruskal–Wallis test.

Effect of LDB1 knockdown in fetal liver cells

We next examined the importance of the LDB1/GATA1 complex in contributing to embryonic erythropoiesis. Fetal liver cells at E12.5 were treated with anti-GFP shRNA (shGFP) as a negative control or anti-LDB1 shRNAs (sh1 and sh2). **Figure 5-a** shows the western blot to show that the levels of LDB1 protein decreased in the fetal liver cells from day 1 (D1) till day 3 (D3) after the knockdown. On day 3, the cells were also FACS sorted based on TER119 and CD71 presence. The P2, P3 and P4 population of cells (see above) decreased when compared to shGFP (**Fig5-b**). This result shows that the LDB1 protein likely plays a key regulatory role at the pro-erythroblast stage of erythropoiesis in the fetal liver.

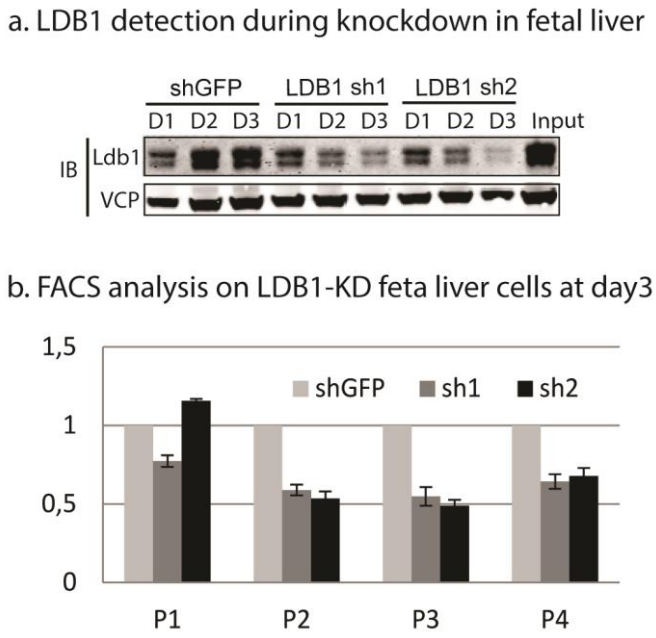


Figure 5. LDB1 KD effects on erythropoiesis

a. Western blot detection of LDB1 knockdown efficiency in fetal liver cells on day1 (D1) to day3 (D3).

b. FACS analysis of LDB1-KD fetal liver cells on day 3. Cell sorting based on TER119 and CD71. Four populations (P1 to P4) are described as Fig5-a. Error-bars present standard error of the mean.

Discussion

Our results show that the LDB1/GATA1 and LDB1/E2A complexes appear during the BL-CFC stage of ES cell differentiation. In the fetal liver at E12.5, we show that the GATA1/LDB1 complex is found at high levels in the proerythroblasts, which reflects the importance of the LDB1 complex in definitive erythropoiesis.

GATA2 is present during the early stage of erythropoiesis, and binds to its own promoter functioning as an activator [17]. It also binds to the *Gata1* gene promoter to activate its expression. *Gata1* expression in turn represses the *Gata2* gene by binding to its promoter via the FOG1/MeCP1 complex, while it activates its own expression. Thus, this "GATA factor switching" represents a forward drive towards the late stage of erythroid differentiation, achieved through changes in gene expression pattern [17, 18]. GATA2 regulates genes that are important for the proliferation of stem or progenitor cells whereas GATA1 provokes the final erythroid fate by regulating the expression of erythroid specific genes. We identified the LDB1/GATA1 and LDB1/E2A complexes to be present already in day-4 ES cell EBs, which suggests that the LDB1 complex may already recruit its components; however, we do not know the function of this complex and whether GATA factors are involved at this stage. From our PLA results, we do not know which cell population contains the LDB1/GATA1 and LDB1/E2A complexes in the day-4 EBs. BL-CFCs appear between ES cell differentiation day 3.75 to day 4.25 showing high levels of vascular endothelial growth factor receptor-2 (VEGFR2, also known as FLK1). The absence of LDB1 results in a decreased number of BL-CFCs in the *Ldb1* KO EBs compared to wild-type EBs and they fail to generate hematopoietic and endothelial lineages [19]. LDB1 binds within or around the gene body of *Gata2*, *Scl*/*Tal1*, *Runx1* and *Gif1b* and these genes are down-regulated in *Ldb1* KO- BL-CFCs, indicating that the LDB1 complex is essential for the activation of specific hematopoietic genes. LDB1 binding sites in the BL-CFCs revealed several DNA-binding motifs, one of which is the prominent known E-box:Gata motif referring the binding of GATA factors at the same sites with LDB1. It was therefore clear that the LDB1 complex is essential for early embryonic hematopoiesis. Our results show that the GATA1/LDB1 interaction already takes place at that stage but do not show whether it is already playing an important role. It has been shown that *Gata1* KO ES cells did not affect the formation of clonogenic progenitors in chimeric *in vitro* differentiation, and *Gata1* KO colonies contained phenotypically normal macrophages, neutrophils and megakaryocytes [20]. These results suggest that GATA1 is still not functional at this early stage. Our LDB1/GATA1 PLA experiments in day4 EBs suggest that this complex may already be recruited to its binding sites and ready for its later function. The E-box:Gata motif may be bound by the GATA2 protein. It would be interesting to identify whether there is also an LDB1 complex containing GATA2 before GATA1 and

whether this is functional, since GATA2 is important for the generation of FLK1+ BL-CFCs at the same stage of *in vitro* ES cell differentiation [21] and *Gata2* KO mouse embryos die at E11.5 with severe anemia [22]. Unfortunately we were not able to successfully perform a GATA2 PLA due to the quality of the GATA2 antibody.

The RING domain-containing E3 ubiquitin ligase RNF12 can bind to LDB1 as a negative repressor and recruit the Sin3A-containing HDAC repression complex [23]. RNF12 ubiquitinates the Lys-134 residue of LDB1 protein to change its stability [24]. RNF12 expression shows a decrease at day 4 ES cell differentiation [25] which may be responsible for the LDB1 RNA level increase we find reaching a peak at day 5 in the whole EBs (data not shown). However both RNF12 and LDB1 are already expressed in ES cells, suggesting there must be another mechanism to protect LDB1 from RNF12-mediated degradation. For instance, we find the single-stranded DNA-binding proteins (SSDBs) also to be present and these have been shown to interact with LDB1 to stabilize its activity [26].

The HLH heterodimer E2A, consisting of E12 and E47 is essential factor for T- and B-cell development. In MEL cells, E2A is binding with LDB1 in the LDB1 complex to regulate its target genes [12]. E2A can recruit the co-activator protein p300/CBP to the promoter of target genes for the cell-type specification [27-29]. The LDB1/E2A complex may function as a (sub)complex during the very early differentiation, or may first have to recruit its multiple binding partners, such as TAL1 and LYL1 to become a functional LDB1 complex. In FLK1+ cells, the LDB1 pull-down followed by mass spectrometry showed the presence of other binding partners, such as TAL1, E2A, LMO2 and SSBP proteins but no GATA factors (unpublished data). We think that one possibility is that the GATA1 protein is not detectable due to its low amount, or the LDB1 complex is still recruiting its components at this stage, and the LDB1 sub-complex may function without GATA factors on its stage-specific target genes.

FOG1 has two isoforms in MEL cells. Only the longer version of FOG1 (FOG1-L) binds MTA2 protein which is a key component of the MeCP1/NuRD complex. Since we used rabbit polyclonal antibody to detect both isoforms of FOG1 protein in the PLA experiments, it is not clear which isoform binds to GATA1 at day 5 of ES cell differentiation. From the size-exclusion-chromatography experiments (supplementary data, **Figure S1**), it is clear that the majority of GATA1 and FOG1 are not in the same fraction at day 4 and day 5, but there is a clear signal in the high MW fractions that show GATA1, FOG1 and MTA2. This is likely responsible for the PLA detection at day 5, particularly because PLA can detect much less protein than western-blotting [30].

In fetal liver the GATA1/LDB1 interaction peaks in the CD71+/TER119- cell population i.e. at a relatively early stage of erythropoietic differentiation. This interaction decreases during the further erythropoiesis. It is known that both the LDB1 and GATA1 proteins can be post-translationally modified. GATA1 binding to other factors is highly influenced by the posttranscriptional modifications of the GATA1

protein of which several have been reported. Sumoylation of GATA1 on Lys-137 within the N-terminal transactivation domain by the SUMO-1 and PIASy proteins, leads to repression of GATA1 transcriptional activity [31]. Erythropoietin signaling promotes the dissociation of HDAC5 from GATA1 [32]. SENP1 (SUMO-specific protease) can deconjugate SUMO from GATA1 to regulate the expression of GATA1-dependent genes [33]. The acetyltransferase p300/CBP can acetylate GATA1 protein at its two zinc-finger domains to regulate GATA1 activity of DNA binding and protein partner interactions [34]. Thus, the modification status of GATA1 protein strongly influences its activation or repression activities together with its binding partners. It would be interesting to know which posttranscriptional modifications of the GATA1 protein are present in the CD71+/TER119- cell population and know the target genes.

I have summarized the LDB1 and GATA1 complexes in **Figure 6**. During early embryogenesis, the GATA2 complex activates the *Gata2* and *Gata1* genes in the hemangioblasts (BL-CFCs). At this stage, the GATA1/LDB1/E2A may start recruiting all its "erythroid" binding partners preparing for its function at later stages, while GATA1/FOG1 complex appears afterwards. The changes of GATA1/LDB1 complex from BL-CFC stage to the erythroid commitment stage are still unknown. GATA2/LDB1 may take place at these stages since GATA2 is important for HSC maintenance [35]. After the erythroid commitment, GATA1/LDB1 complex increases until the proerythroblast stage and decreases when erythroid maturation started. GATA1/FOG1 together with MeCP1 can repress early-differentiation programme genes including *Gata2* to achieve erythroid commitment. Once the development of the erythroid lineage is in progress, GATA1/LDB1 activates erythroid specific genes such as *Alas2* and *Gypa* etc.

In summary, our study first demonstrates the LDB1 complex increases from early precursor stages (CD71-/TER119-) to proerythroblasts (CD71+/TER119-) in the fetal liver, showing a peak at the early stage of erythroblast maturation. Our results also showed the sequential emergence of GATA1 complexes, GATA1/LDB1 and GATA1/FOG1, which happens during the *in vitro* ES cell differentiation at the time point of the emergence of BL-CFCs.

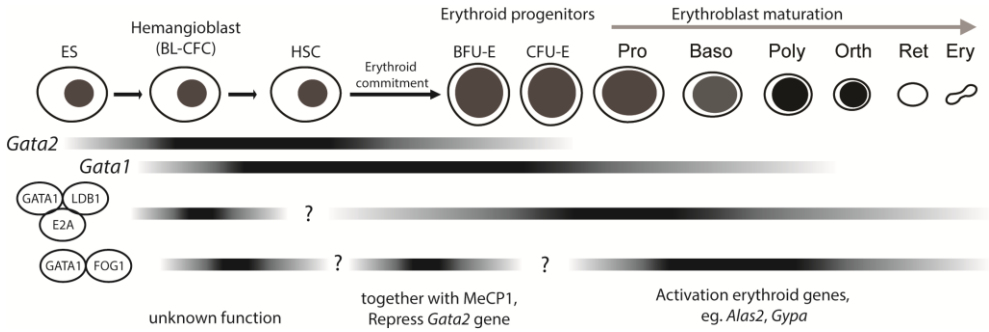


Figure 6. Schematic representation of GATA complex during the embryonic development
Gata2 and *Gata1* expression, together with GATA1/LDB1/E2A and GATA1/FOG1 complexes are shown as gray bars. Question marks indicate unknown changes of the indicated complex. Texts below the gray bars indicate the complex function at different stages.
Abbreviations: ES, embryonic stem cell; BL-CFC, blast colony-forming cell; HSC, hematopoietic stem cell; BFU-E, burst forming unit-erythroid; CFU-E, colony forming unit-erythroid; Pro, immature proerythroblast; Baso, basophilic erythroblast; Poly, polychromatic erythroblast; Orth, orthochromatic erythroblast; Ret, reticulocytes; Ery, erythrocyte.

Material and Methods

ES cell line and culture

Embryonic stem cells were cultured in DMEM (BE12-604F/U1, Lonza) containing 15% FCS, 1% non-essential amino acids (BE13-114E, Lonza), 100units/ml penicillin and 100mg/ml streptomycin (GIBCO/BRL), 6.3e-4% 2-mercaptoethanol (115433, Merck), 100U/ml Esgro (LIF) (ESG1106, Millipore).

ES cell differentiation by the hanging drop method

For ES cell differentiation, cells were detached and suspended in IMDM (31980, Gibco) containing 15% FCS, 1% non-essential amino acids (BE13-114E, Lonza), 100units/ml penicillin and 100mg/ml streptomycin (GIBCO/BRL), 0.05mg/ml Ascorbic Acid (A0278-25G, Sigma), 0.3mg/ml Human Transferrin (Product code: T101-5, SCIPAC), 2.34e-5% 1-thioglycerol (M6145-25ML, Sigma), 5% PFHM-II (12040, Gibco). Two hundred cells per drop in 15ul medium were used to make a hanging drop on the inner face of a 10cm dish lid with HBSS (24020, Gibco) in the dish to maintain humidity. Embryoid bodies (EBs) were collected by flushing with 40ml 1xPBS and waiting until the EBs sank down by gravity in a 50ml falcon tubes.

RNA extraction and reverse transcriptase PCR

EBs were collected on different days and dissociated by Trypsin-EDTA for five minutes, after which FCS was added to a final concentration of 1%. After 1xPBS wash, RNA was isolated using Trizol (Invitrogen). cDNA was synthesized using the SuperScript® III First Strand Synthesis System and random hexamer primers (Invitrogen). Real-time PCR was performed using Platinum Taq Polymerase and SyberGreen (Invitrogen) on a Bio-Rad CFX96 PCR system. Ribonuclease/angiogenin inhibitor 1 (Rnh1) mRNA was amplified in parallel for normalization. Primer sequences are listed below.

Rnh1-F TGCAGGCACTGAAGCACCA, Rnh1-R TCCAGTGTGAGCAGCTGAG;
LDB1-F GTGACAATCTCTGGTGGGA, LDB1-R GGAAGTAGCGTGGTATCAG;
GATA1-F TGCCTGTGGCTTGTATCA, GATA1-R TGTTGTAGTGGTCGTTTGAC;
LMO2-F CGAAAGGAAGAGCCTGGAC, LMO2-R CCCTATGTTCTGCTGGCA;
Gata2-F (2007)* GCAGAGAAGCAAGGCTCGC,
Gata2-R (2007)* CAGTTGACACACTCCCGGC; *Gata2 primer set is from Ref [36].

Antibodies

The following antibodies were used in the western blots: anti-LDB1 (sc-11198), GATA1 (sc-1234 or sc-265), Tal1 (sc-22809), E2A (sc-349) from Santa Cruz and LMO2 (ab72841) from Abcam. For FACS we used CD71-FITC (553266), Ter119-PE (553673)

from BD Pharmingen, Rabbit anti-CD71 (sc-9099, Santa Curz), anti-Rabbit A647 (A21245, Invitrogen).

Nuclear extract preparation and Immunoprecipitation

Nuclear extract was prepared as described by de Boer et al. (de Boer et al., 2003) Briefly, cells were lysed in BufferA (10mM HEPES-KOH pH7.9, 1.5mM MgCl2, 10mM KCl, 0.5mM DTT, 1x Protease inhibitor) for 10 minutes, followed by incubation in BufferB (20mM HEPES-KOH pH7.9, 25% Glycerol, 420mM NaCl, 1.5mM MgCl2, 0.2mM EDTA, 0.5mM DTT, 1x Protease inhibitor) for 30 minutes. The protein concentration was determined using Pierce® BCA Protein Assay (Thermo). 500µg total protein extract was used for each IP and incubated in Heng150 buffer (20mM HEPES-KOH pH7.9, 20% Glycerol, 0.25mM EDTA pH8, 0.05% NP40, 150mM KCl, 1x protease inhibitor) over-night at $4C^0$. Protein A/G PLUS-Agarose beads (Santa Cruz Biotechnology, sc-2003) were blocked with CEA for one hour at room temperature and added to the IP samples for another hour in $4C^0$. After washing with Heng150 buffer three times, 30µl Heng150 buffer was added to suspend the IP samples.

Western Blots

Samples in Laemmli buffer were boiled on a heat block for five minutes and loaded on NuPAGE precast 4-12% gradient Bis-Tris acrylamide gels (Novex). Proteins were transferred onto nitrocellulose transfer membrane (Whatman) and blocked with 1% fat-free skim milk (70166-500G, SIGMA) in PBS-Tween. Detection of specific proteins was using the antibodies as specified. Fluorescent secondary antibodies were used for visualization by the Odyssey system (LI-COR).

Mass Spectrometry

LDB1 IPs in P2 and P3 sorted fetal liver-cell populations were digested and dimethyl labeled as described by de Boer [37]. Samples were processed and analyzed by mass spectrometry (LS-MSMS, MS) as described previously REF [13].

RNA interference

Lentiviral shRNAs for LDB1 were described previously by Stadhouders (REF [38]).

Proximity ligation assay on EB and mouse embryo tissue

At day 4, day 5 and day 9, EBs were collected and fixed with 2% PFA in PBS for one hour. Followed by 15% Sucrose incubation for 30 minutes, EBs were frozen in OCT Tissue-Tek (Sakura) and sliced into 10µm tissue on slides. Mouse embryos (embryonic day 12.5) were treated the same but fixed after slicing. 0.1% Triton-X (Cat. T8787, Sigma) in PBS was used to permeablize EB or embryo-tissue slices. Primary antibodies were incubated over-night at $4C^0$ in dilutions of 1⁄200 for anti-LDB1 and 1⁄500 for

anti-GATA1 antibodies. Secondary antibody PLA probes (OLINK) were incubated for an hour at 37C^{o}. By adding oligonucleotides and DNA ligase, a rolling circle was formed and amplified by polymerase for 100 minutes at 37 degree. The fluorescent detection probes were added after the amplification. Slides were embedded in VectaSheild (Olink) and pictures were taken under a Leica SP5 confocal microscope. PLA signals were analyzed in BlobFinder software (Uppsala University). To compare the signal in different groups, the experiments were performed on three distinct days. The Kruskal–Wallis test for variance between groups was performed and the Bonferroni method to counteract multiple comparison errors was applied.

Flowcytometry analysis and Cell sorting

Mouse fetal liver cells were incubated with CD71-FITC and Ter119-PE antibodies in PBS containing 1% bovine serum albumin for 30 minutes. The cells were washed and Hoechst (Invitrogen) was added to a final concentration of 1⁄5000. Cells were analyzed and sorted on a FACSAria III (Becton Dickinson). For staining of LDB1 shRNA treated cells, Rabbit anti-CD71 and anti-rabbit AF647 antibodies were used.

Size-Exclusion Chromatography

Nuclear extracts from MEL, Day 4 and Day 5 EBs are loaded on an AKTA FPLC apparatus with a Superose 6 10⁄30 column (Amersham Biosciences). Each fraction contains 500ul flow-through. Fractions were precipitated with trichloroacetic acid and analyzed by Western blotting.

Supplementary data

Figure S1. GATA1 and LDB1 form distinct complexes during ES cell differentiation

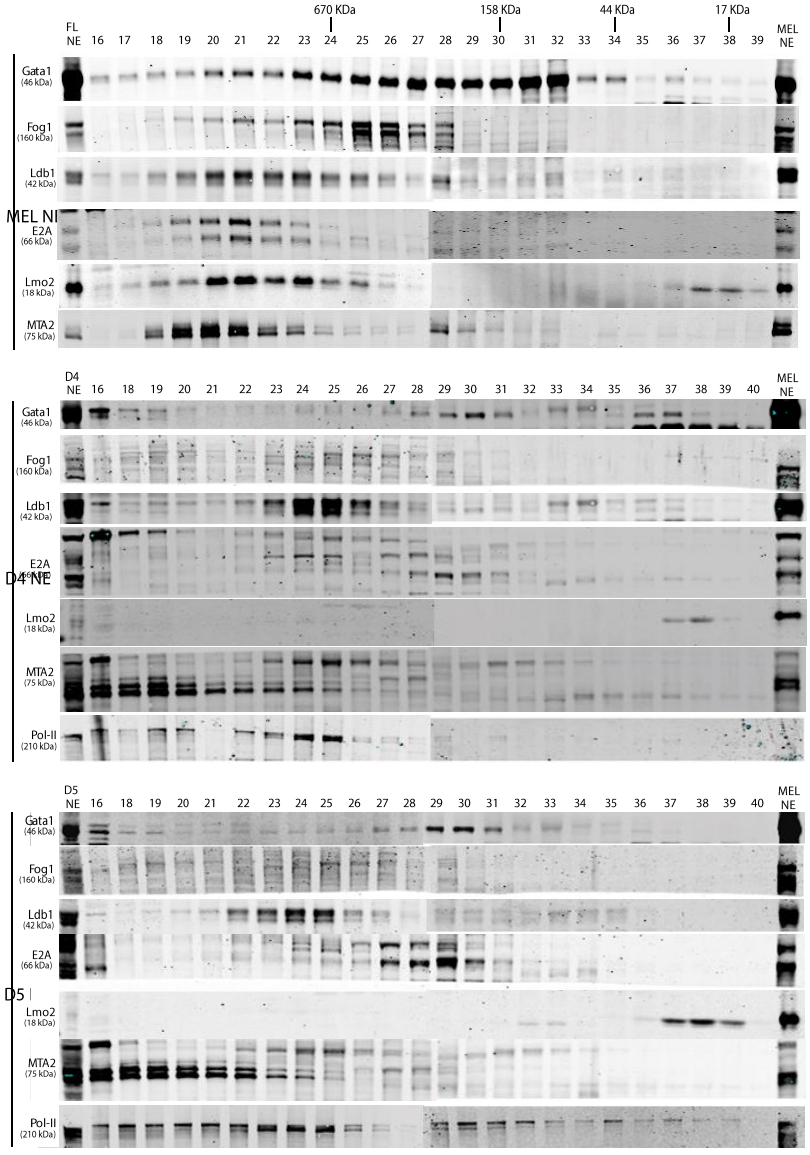


Figure S1. GATA1 and LDB1 form distinct complexes during ES cell differentiation
Native complexes were separated from day 4 EB, day 5 EB and MEL cell nuclear extracts using Superose 6 column (Each lane presents the fraction of 500ul fraction). Larger to smaller molecular mass fractions are shown from left to right. The sizes of complexes are described above. The last lane is the MEL extracts as a positive control.

Table S1. LDB1 and FOG1 binding partners on Day 4 and Day 5 ES cell differentiation

*numbers indicate the mascot score

	LDB1-IP MS		FOG1-IP MS	
	D4 NE	**D5 NE**	**D4 NE**	**D5 NE**
IP protein	LDB1(770)	LDB1(725)	FOG1(614)	FOG1(741)
LDB1 complex	LHX1(595), SSBP2(335), SSBP3(463), SSBP4(311), LMO4(212), LDB2(61)	LHX1(137), SSBP1(129), SSBP2(408), SSBP3(409), LMO4(190), LDB2(155), TAL1(55)	-	-
NuRD complex	-	HDAC2(296), MTA2(101), MTA3(56), CHD4(791), MBD3(101)	HDAC1(670) HDAC2(569) MTA1(709) MTA2(1079) MTA3(578) KDM1a(271) CHD4(3461) MBD3(447) GATAD2a (808) GATAD2b (570)	HDAC1(566), HDAC2(460), MTA1(751), MTA2(816), CHD4(2593), GATAD2a(351)
SIN3 HDAC complex	-	HDAC2(296), RBBP7(258), SIN3a(215)	HDAC1(670) HDAC2(569) SIN3a(366) RBBP7(521)	HDAC1(566), HDAC2(460), SIN3a(243), RBBP7(520)
BHC complex	-	-	HDAC1(670), HDAC2(569), ZMYM2(169), RCOR2(144)	HDAC1(566), HDAC2(460), ZMYM2(50), RCOR2(117)
Mediator complex	-	-	MED1(67), MED4(382), MED12(162), MED13(141), MED15(160), MED16(71), MED17(63), MED22(48), MED23(95), MED24(130), MED27(135), MED30(72), MED31(115)	MED15(120), MED17(86), MED27(69)
SWI/SNF complex	-	-	SMARCB1(186), SMARCD1(302), SMARCC1(879), SMARCC2(385), SMARCE1(188), ARID1a(161), ACTL6a(61), ACTL6b(61), SMARCA4(1081), PBRM1(421)	SMARCD1 (286), SMARCC1 (455), SMARCC2 (169), SMARCA2(256), SMARCA4(358), PBRM1(419)
Other proteins	ILF2 (266), ILF3 (214), TCEA1 (125), TCERG1 (238), CDK11b (85), SMARCA5 (208), KDM1a (269)	TCEB1(221), TCERG1(55), LMX1b(71), NCOR1(174), EIF4g1(810), EIF4a2(566), Ik(251), CDK11b(150), ISL1(244), RNF2(83)	NCOR2(67), IGF2bp2(217), ILF2(322), ILF3(200), TCERG1(376), CDK9(170), NCOA3(712), EP300(447), NAT10(238), RBM15(110), POU5f1(187)	NCOR2(68), HMGA2(104), CREBBP(57), TCERG1(64), IK(294), NCOA3(197), EIF4g1(461), CDK1(236), CDK12(174), TCEB1(134), KDM1a(245)

Table S2. Top 15 categories of IPA bio-functional analysis on gene lists from LDB1-IP and FOG1-IP mass spectrometry

Day 4 LDB1-IP proteins:

	Category	p-value
1	RNA Post-Transcriptional Modification	6.8E-14-2.53E-02
2	Protein Synthesis	1.4E-09-2.05E-02
3	Gene Expression	1.18E-08-2.73E-02
4	Cellular Assembly and Organization	1.19E-05-2.53E-02
5	Cellular Function and Maintenance	1.19E-05-2.73E-02
6	DNA Replication, Recombination, and Repair	3.16E-05-2.53E-02
7	Cell Cycle	8.39E-05-2.73E-02
8	Cell-To-Cell Signaling and Interaction	8.39E-05-2.71E-02
9	Cellular Movement	8.39E-05-2.73E-02
10	Tissue Development	8.39E-05-2.73E-02
11	Molecular Transport	2.42E-04-2.73E-02
12	RNA Trafficking	2.42E-04-4.7E-03
13	Embryonic Development	6.82E-04-2.73E-02
14	Cell Death and Survival	8.03E-04-2.73E-02
15	Organismal Development	8.24E-04-2.73E-02

Day 5 LDB1-IP proteins:

	Category	p-value
1	RNA Post-Transcriptional Modification	1.01E-07-1.16E-02
2	Protein Synthesis	4.25E-07-1.83E-02
3	Cellular Assembly and Organization	6.15E-06-2.31E-02
4	Cellular Function and Maintenance	6.15E-06-2.23E-02
5	Cell Cycle	7.71E-06-2.31E-02
6	Gene Expression	8.32E-06-2.04E-02
7	Hematological System Development and Function	1.1E-05-2.31E-02
8	Cellular Movement	1.39E-05-2.31E-02
9	Molecular Transport	1.76E-05-1.16E-02
10	Protein Trafficking	1.76E-05-1.76E-05
11	Cellular Growth and Proliferation	3.39E-05-2.31E-02
12	Cell Death and Survival	7.2E-05-2.31E-02
13	Organismal Survival	7.77E-05-5.38E-04
14	Skeletal and Muscular System Development and Function	9.12E-05-2.31E-02
15	Post-Translational Modification	9.89E-05-1.45E-02

Day 4 FOG1-IP proteins

	Category	p-value
1	RNA Post-Transcriptional Modification	2.39E-27-1.05E-02
2	Gene Expression	3.76E-13-2.2E-02
3	Cell Cycle	1.79E-11-2.2E-02
4	DNA Replication, Recombination, and Repair	1.79E-11-2.2E-02
5	Cellular Assembly and Organization	1.7E-10-2.2E-02
6	Cell Signaling	4.29E-10-2.2E-02
7	Cellular Compromise	2.29E-08-2.2E-02
8	Cellular Movement	5.81E-08-2.2E-02
9	Protein Synthesis	8.71E-07-1.57E-02
10	Molecular Transport	2.09E-06-2.2E-02
11	RNA Trafficking	2.09E-06-4.63E-03
12	Embryonic Development	1.48E-05-2.2E-02
13	Organismal Survival	1.48E-05-1.48E-05
14	Organismal Development	4.14E-05-2.2E-02
15	RNA Damage and Repair	5.95E-05-5.95E-05

Day 5 FOG1-IP proteins

	Category	p-value
1	Cell Cycle	7.14E-11-1.18E-02
2	Cellular Assembly and Organization	3.61E-10-1.18E-02
3	RNA Post-Transcriptional Modification	6.02E-10-1E-02
4	Cell-To-Cell Signaling and Interaction	2.51E-08-1.18E-02
5	Tissue Development	2.51E-08-1.18E-02
6	Gene Expression	1.76E-07-1.18E-02
7	Cell Death and Survival	1.62E-06-1.18E-02
8	Skeletal and Muscular System Development and Function	3.89E-06-1.17E-02
9	Cell Morphology	5.12E-06-1.18E-02
10	Embryonic Development	5.74E-06-1.18E-02
11	Organismal Development	5.74E-06-1.17E-02
12	Connective Tissue Development and Function	5.93E-06-1.18E-02
13	Organismal Survival	1.4E-05-4.55E-03
14	Reproductive System Development and Function	3.12E-05-1.16E-02
15	Organ Morphology	5.54E-05-1.18E-02

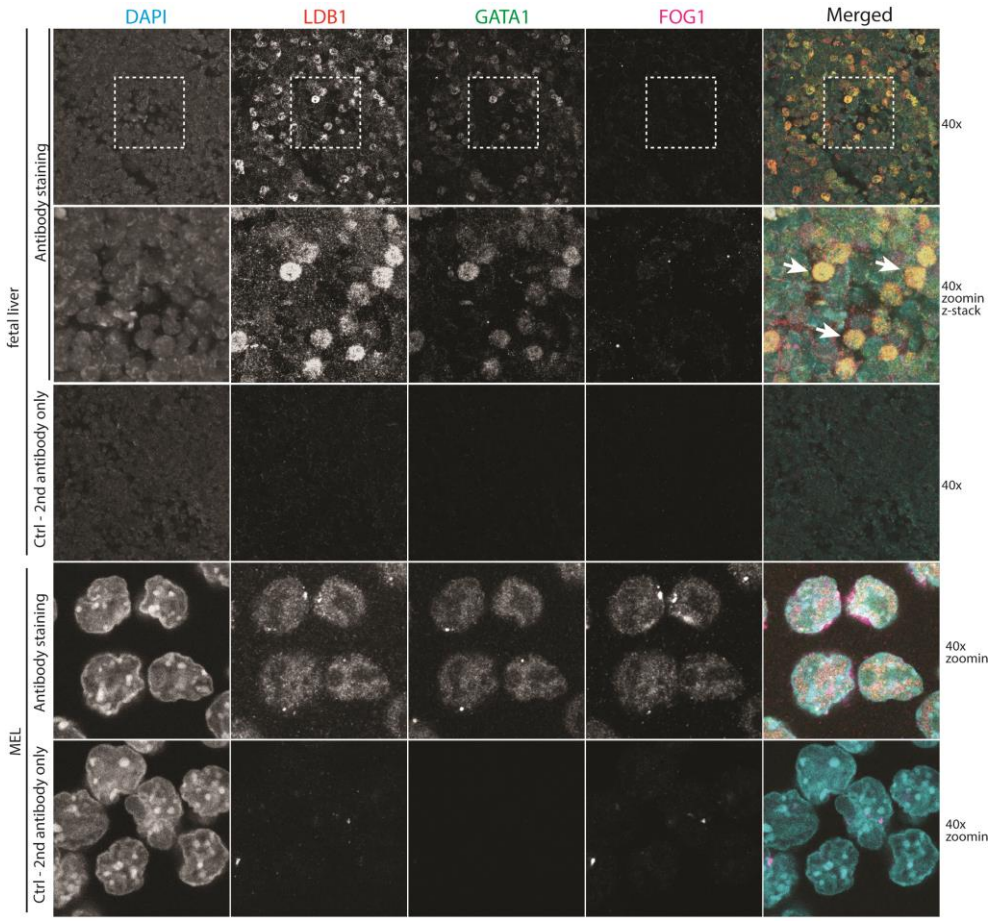


Figure S2. Immunofluroreсence on E12.5 fetal liver tissue

Immunofluorecence for LDB1 (in red), GATA1 (in green) and FOG1 (in purple) on E12.5 fetal liver tissue. DAPI (cyan) stains the nucleus. Single staining pictures from distinct fluorescent channels are indicated in gray. Positive staining controls was performed on MEL cells. The negative secondary antibody alone controls were performed at the same time both on fetal liver and MEL cells. White arrows in merged pictures point out the yellow cells which show co-localization of LDB1 and GATA1 proteins in fetal liver.

References

1. Dzierzak, E. and N.A. Speck, *Of lineage and legacy: the development of mammalian hematopoietic stem cells.* Nat Immunol, 2008. **9**(2): p. 129-36.
2. Medvinsky, A., S. Rybtsov, and S. Taoudi, *Embryonic origin of the adult hematopoietic system: advances and questions.* Development, 2011. **138**(6): p. 1017-31.
3. Snow, J.W. and S.H. Orkin, *Translational isoforms of FOG1 regulate GATA1-interacting complexes.* J Biol Chem, 2009. **284**(43): p. 29310-9.
4. Mukhopadhyay, M., et al., *Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation.* Development, 2003. **130**(3): p. 495-505.
5. Fredriksson, S., et al., *Protein detection using proximity-dependent DNA ligation assays.* Nat Biotechnol, 2002. **20**(5): p. 473-7.
6. Kennedy, M., et al., *A common precursor for primitive erythropoiesis and definitive haematopoiesis.* Nature, 1997. **386**(6624): p. 488-93.
7. Choi, K., et al., *A common precursor for hematopoietic and endothelial cells.* Development, 1998. **125**(4): p. 725-32.
8. Lancrin, C., et al., *Blood cell generation from the hemangioblast.* J Mol Med (Berl), 2010. **88**(2): p. 167-72.
9. Shawlot, W., et al., *Lim1 is required in both primitive streak-derived tissues and visceral endoderm for head formation in the mouse.* Development, 1999. **126**(22): p. 4925-32.
10. Bally-Cuif, L. and E. Boncinelli, *Transcription factors and head formation in vertebrates.* Bioessays, 1997. **19**(2): p. 127-35.
11. Allalou, A. and C. Wahlby, *BlobFinder, a tool for fluorescence microscopy image cytometry.* Comput Methods Programs Biomed, 2009. **94**(1): p. 58-65.
12. Meier, N., et al., *Novel binding partners of Ldb1 are required for haematopoietic development.* Development, 2006. **133**(24): p. 4913-23.
13. Rodriguez, P., et al., *GATA-1 forms distinct activating and repressive complexes in erythroid cells.* EMBO J, 2005. **24**(13): p. 2354-66.
14. Anguita, E., et al., *Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2.* EMBO J, 2004. **23**(14): p. 2841-52.
15. Goardon, N., et al., *ETO2 coordinates cellular proliferation and differentiation during erythropoiesis.* EMBO J, 2006. **25**(2): p. 357-66.
16. Hu, X., et al., *LSD1-mediated epigenetic modification is required for TAL1 function and hematopoiesis.* Proc Natl Acad Sci U S A, 2009. **106**(25): p. 10141-6.
17. Grass, J.A., et al., *GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8811-6.
18. Kaneko, H., R. Shimizu, and M. Yamamoto, *GATA factor switching during erythroid differentiation.* Curr Opin Hematol, 2010. **17**(3): p. 163-8.
19. Mylona, A., et al., *Genome-wide analysis shows that Ldb1 controls essential hematopoietic genes/pathways in mouse early development and reveals novel players in hematopoiesis.* Blood, 2013. **121**(15): p. 2902-13.
20. Pevny, L., et al., *Development of hematopoietic cells lacking transcription factor GATA-1.* Development, 1995. **121**(1): p. 163-72.
21. Lugus, J.J., et al., *GATA2 functions at multiple steps in hemangioblast development and differentiation.* Development, 2007. **134**(2): p. 393-405.
22. Tsai, F.Y., et al., *An early haematopoietic defect in mice lacking the transcription factor GATA-2.* Nature, 1994. **371**(6494): p. 221-6.
23. Ostendorff, H.P., et al., *Ubiquitination-dependent cofactor exchange on LIM homeodomain transcription factors.* Nature, 2002. **416**(6876): p. 99-103.
24. Howard, P.W., et al., *Regulation of LIM-domain-binding 1 protein expression by ubiquitination of Lys134.* Biochem J, 2010. **429**(1): p. 127-36.
25. Zhang, L., et al., *RNF12 controls embryonic stem cell fate and morphogenesis in zebrafish embryos by targeting Smad7 for degradation.* Mol Cell, 2012. **46**(5): p. 650-61.
26. Xu, Z., et al., *Single-stranded DNA-binding proteins regulate the abundance of LIM domain*

and LIM domain-binding proteins. Genes Dev, 2007. **21**(8): p. 942-55.
27. Kee, B.L., J. Arias, and M.R. Montminy, *Adaptor-mediated recruitment of RNA polymerase II to a signal-dependent activator.* J Biol Chem, 1996. **271**(5): p. 2373-5.
28. Massari, M.E. and C. Murre, *Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms.* Mol Cell Biol, 2000. **20**(2): p. 429-40.
29. Bayly, R., et al., *Critical role for a single leucine residue in leukemia induction by E2A-PBX1.* Mol Cell Biol, 2006. **26**(17): p. 6442-52.
30. Gustafsdottir, S.M., et al., *Proximity ligation assays for sensitive and specific protein analyses.* Anal Biochem, 2005. **345**(1): p. 2-9.
31. Collavin, L., et al., *Modification of the erythroid transcription factor GATA-1 by SUMO-1.* Proc Natl Acad Sci U S A, 2004. **101**(24): p. 8870-5.
32. Delehanty, L.L., G.C. Bullock, and A.N. Goldfarb, *Protein kinase D-HDAC5 signaling regulates erythropoiesis and contributes to erythropoietin cross-talk with GATA1.* Blood, 2012. **120**(20): p. 4219-28.
33. Yu, L., et al., *SENP1-mediated GATA1 deSUMOylation is critical for definitive erythropoiesis.* J Exp Med, 2010. **207**(6): p. 1183-95.
34. Boyes, J., et al., *Regulation of activity of the transcription factor GATA-1 by acetylation.* Nature, 1998. **396**(6711): p. 594-8.
35. Li, L., et al., *Nuclear adaptor Ldb1 regulates a transcriptional program essential for the maintenance of hematopoietic stem cells.* Nat Immunol, 2011. **12**(2): p. 129-36.
36. Holley, M.C., A. Kneebone, and M. Milo, *Information for gene networks in inner ear development: a study centered on the transcription factor gata2.* Hear Res, 2007. **227**(1-2): p. 32-40.
37. de Boer, E., et al., *Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice.* Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7480-5.
38. Stadhouders, R., et al., *Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development.* EMBO J, 2012. **31**(4): p. 986-99.

Chapter 3

CDC73 is a dual function co-regulator of the erythroid LDB1 complex

Manuscript in preparation

CDC73 is a dual function co-regulator of the erythroid LDB1 complex

Abstract

The *Lmo2* gene is an important regulator of hematopoiesis and its misregulation is associated with human leukemia. LMO2 overexpression can also impair erythroid differentiation. In erythroid cells, LMO2 is part of a multiprotein complex containing the essential transcription factors (TFs) GATA1/LDB1/TAL1, referred to as the LDB1 complex, which is a key regulator of erythroid genes. However, the exact roles of LMO2 in the GATA1/LDB1/TAL1 complex are unknown. In the present study, we show that LMO2 binds to CDC73, a component of PAF1 complex, and regulates cell proliferation related genes in MEL cells. These results highlight the dual function of CDC73 with LDB1 and LMO2, which can maintain erythroid cells in the premature stages and, in conjunction with the PAF1 complex, also regulate cell proliferation.

Introduction

The LIM domain only protein LMO2 is an essential factor functioning as a scaffold protein in the GATA1/LDB1/TAL1 complex which contains GATA1, LDB1, TAL1, E2A, ETO2, LSD1 and LMO2 transcription factors [1]. This complex is essential in erythroid differentiation. Loss of *lmo2* in zebrafish impairs primitive erythropoiesis. In mice it is embryonic lethal with absence of blood islands [2, 3]. In human, inappropriate regulation of *Lmo2* frequently leads to T-cell acute lymphoblastic leukemia [4]. Overexpression of LMO2 in fetal liver decreases erythroid development, suggesting a negative regulatory role for LMO2 in erythropoiesis [5]. Visvader and colleagues suggested that the LDB1/LMO2 complex maintains the erythroid precursors in a immature state [6]. However, overexpression of mutants in the LIM2 domain of LMO2 inhibits erythroid differentiation [5], suggesting a positive role for LMO2. In addition it has been reported that enforced expression of LMO2 in FDCP-mix cells enhanced erythroid differentiation [7]. These results suggest that it may have a dual role.

Transcription factors like GATA1, LDB1 and TAL1 are important regulators of the erythroid transcriptiome. Interestingly, the cyclin-dependent kinase 9 (CDK9), a subunit of the P-TEFb complex which functions as a critical positive factor to switch the RNA Pol II from a pausing stage to an actively elongating stage [8-10], has been identified as bona fide component of this complex [1]. The GATA1/LDB1 complex appears to recruit the P-TEFb complex (CDK9) to activate its target genes in erythropoiesis via long range interactions [11].

It has been shown that LDB1 can form different complexes with GATA1, TAL1, E2A, LMO2 and LMO4 [1]. The two LIM-only proteins, LMO2 and LMO4, behave differently in the LDB1 complexes. The protein level of LMO2 in the LDB1 complex stays constant during MEL cell differentiation, while the LMO4 level increases. However, the distinct function of these proteins in the LDB1 complexes is still unknown. In order to answer this question for LMO2, we determined the co-factors of LMO2 by mass spectrometry (LC-MSMS) in MEL cells. Interestingly we found CDC73, a protein that belongs to the PAF1 (RNA polymerase-II associated factor 1) complex regulating the transcription cycle. Using ChIP sequencing and gene ontology analysis, we found that CDC73, together with LMO2 and LDB1, negatively regulates the proliferation and transcription of red-blood-cell related genes in MEL cells. We suggest that the GATA1/LDB1 complex acts as a positive transcription factor complex linking to the polymerase machinery via LMO2 and CDC73 to the PAF1 complex, while it can also form a complex with CDC73 alone to regulate cell proliferation genes in the early stage of erythropoiesis.

Results

Generation of bio-LMO2 and LMO2-bio cells

In order to identify LMO2 interacting partners, we fused a biotinylation tag DNA sequence either to the N- or C-terminus of the LMO2 coding sequence, and stably expressed these constructs in C88BirA MEL cells which express BirA protein-biotin ligase [12]. C88BirA/bioLmo2 or C88BirA/Lmo2bio expressing clones were isolated and selected based on the expression of the fusion proteins. We chose the optimal clones which have an expression level of bio-tagged LMO2 similar to that of endogenous LMO2 as determined by western blotting (**Figure S1**).

Proteomics identification of LMO2 binding partners

Nuclear extracts from C88BirA/bioLmo2 or C88BirA/Lmo2bio MEL cell clones were incubated with streptavidin-coated paramagnetic beads to capture LMO2-containing complexes. The captured proteins were trypsin digested and analyzed by mass spectrometry (LC-MSMS). **Table 1** shows a selection of the captured proteins from both clones (for complete list, see **Table S1**). As expected, members of the LDB1 complex [1] were readily identified such as the LIM-domain binding factor LDB1, the zinc finger protein GATA1, the basic helix-loop-helix proteins TAL1, E2A(TCF3), E2-2(TCF4), HEB(TCF12), and LYL1, together with the ETO family proteins ETO2(CBFA2T3) and MTGR1(CBFA2T2) which are transcriptional repressors and the SSBP proteins which are known binding partners of LDB1 that protect LDB1 from proteosomal degradation [13]. In addition there is CDK9, which together with CCNT1 can phosphorylate Ser-2 in the RNA polymerase C-terminal domain and the transcriptional intermediary factor TIF1-γ which also has an essential role in regulating transcriptional elongation of erythroid genes [14]. Other proteins previously associated with the LDB1 complex like RUNX1 and MYEF2 are also found [15]. MYEF2 was recently identified as a new binding partner of RUNX1 and its knockdown resulted in increased expression of *Runx1*, *Eto2* and *Gata1*, suggesting that MYEF2 functions as

Table 1. LMO2 and CDC73 binding partners in MEL cells

* numbers indicate the mascot score

	LMO2 MS	CDC73 MS
LDB1 complex	LDB1(337), GATA1(106), LMO2(139), TAL1(667), TCF3(566), TCF4(343), TCF12(665), CBFA2T3(417), SSBP2(251), SSBP3(262), SSBP4(232), CBFA2T2(203), RUNX1(110), CDK9(49), LYL1(170), TRIM33(255), MYEF2(65)	LDB1(169), GATA1(54), LMO2(75), TAL1(200), TCF3(281), TCF12(269), CBFA2T3(285), SSBP2(81), SSBP3(114), SSBP4(127), CDK9(350)
PAF1 complex	CDC73(77)	CDC73(2032), CTR9(3788), PAF1(2273), EO1(1895), WDR61(672)
NuRD complex	HDAC1(78), MTA2(46), LSD1(44)	HDAC2(55)

a negative regulator of these genes. Finally, the nucleosome remodeling and histone deacetylase (NuRD) components, HDACs and MTAs were also found. The MTA2 protein has been reported to interact with a longer version of FOG1, which is in one of the GATA1 complexes, repressing GATA1 target genes in the erythroid cells [16]. Based on previous results of GATA1 and LDB1 complexes finding all of these proteins was not unexpected.

However, we also found the CDC73 protein as a binding partner of LMO2. CDC73 (cell division cycle 73, a gene product of HRPT2 locus) is a component of the PAF1 (polymerase-II associated factor) complex that includes PAF1, CDC73, CTR9, LEO1 and RTF1 [17]. In zebrafish, morpholino knockdown of CDC73 can rescue globin expression in *tif1-γ* mutant, suggesting that *cdc73* functions as a suppressor [18]. Chromatin immunoprecipitation (ChIP) on Chip experiments for TIF1-γ and CDC73 using K562 cells revealed that these proteins share common target genes such as *Gata1* and *Gata2* [18]. These results suggest that LMO2 and CDC73 may provide a "missing" link between the LDB1 complex and the RNA polymerase II machinery at the initiation and elongation steps of transcription.

Proteomics identification of CDC73 binding partners

To confirm these results and further investigate the CDC73 complexes, we performed the reciprocal experiment by immunoprecipitation (IP) of the endogenous CDC73 protein from MEL cells followed by mass spectrometry analysis. As also shown in **Table 1**, the CDC73 IP pulled down the core LDB1 complex components GATA1, LMO2, TAL1, TCF3/12, ETO2 and the SSBP proteins. CDK9, essential for the switch of the initiating to the elongating RNA pol II, was also found. Importantly almost all of the components of PAF1 complex (except the RTF1 protein) were found, strongly suggesting that CDC73 is a bona fide component of the PAF1 complex in MEL cells.

To confirm the interaction between LMO2 and CDC73 which was found in the LC-MSMS analysis, immunoprecipitations for these proteins were performed in MEL cells and CDC73, LDB1 and LMO2 proteins were detected by western blots. As shown in **Figure S2**, CDC73 interacts with LMO2 in MEL cells. The negative control in HEK cells does not detect any LMO2. However, in LMO2-IP samples, the detection of CDC73 is almost the same as IgG background (**Figure S2**, lower panel), whereas IP for LMO2 is efficient in MEL-, MEL+ and K562 cells (**Figure S2**, upper panel). The amount of LDB1 is much higher in the LMO2-IP than in CDC73-IP, suggesting that LMO2 and LDB1 directly interact, while LMO2 bridges LDB1 and CDC73. Finally, the CDC73-IP in HEK cells also brings down LDB1 protein and may suggest that LDB1 and CDC73 may form a complex in non-hematopoietic cells.

We conclude that CDC73 binds to LMO2 in the LDB1 complex; other CDC73 complexes may have other functions.

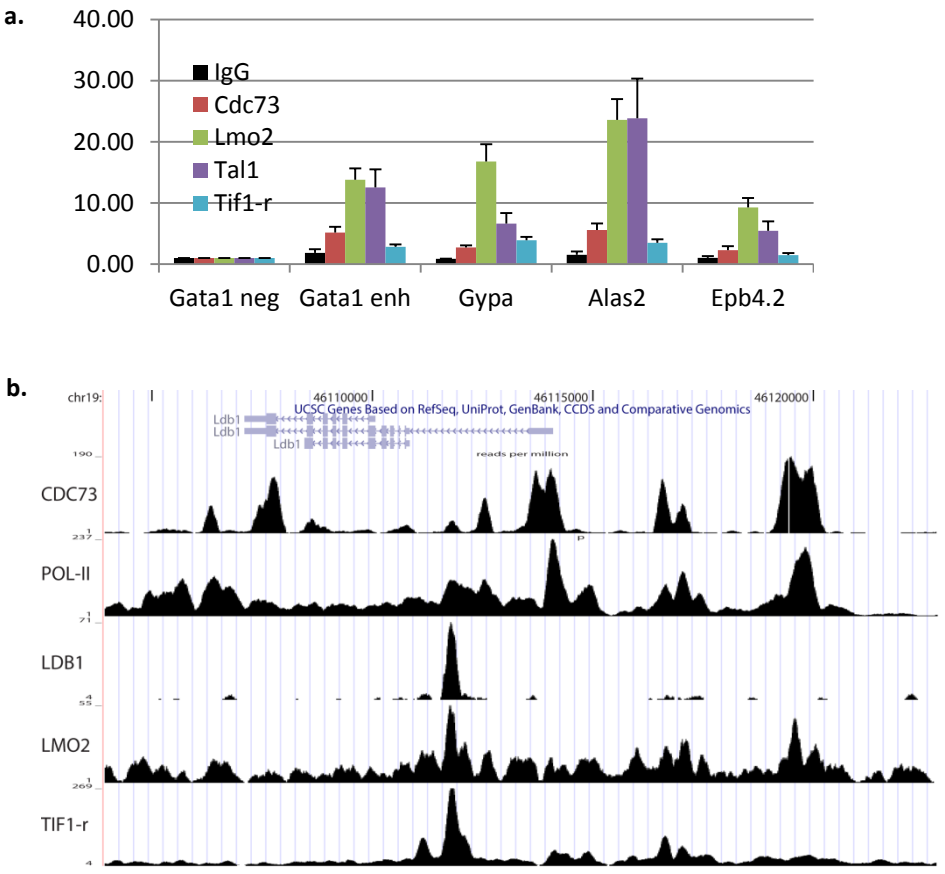


Figure 1. ChIP-qPCR and the peak view on *Ldb1* gene in MEL non-induced cells
a. ChIP-qPCR for CDC73, LMO2 TAL1 and TIF1-γ on *Gata1* negative control region, the *Gata1* enhancer region, and the *Gypa*, *Alas2* and *Epb4.2* gene promoters.
b. ChIP peaks in *Ldb1* gene locus for transcription factors indicated on the left.

Genome-wide CDC73 DNA binding sites and target genes in MEL cells

As a component of the PAF1 and LMO2 complexes, the genome-wide DNA binding sites of CDC73 were identified by ChIP followed by high-throughput sequencing (ChIP-Seq) and compared to the binding sites for RNA Poly II, LMO2 and LDB1. The *Gata1* -3.5kb enhancer region was used as a positive binding site control; while TAL1 ChIP as a known positive control for the ChIP (**Figure 1-a**). Several other erythroid loci, known to be bound by the LDB1 complex were also tested (**Figure 1-a)**. **Figure 1-b** shows the ChIP-Seq peaks at the *Ldb1* gene locus for the transcription factors, CDC73, POL-II, LDB1 and LMO2. CDC73 peaks showed a binding pattern similar to that of RNA Poly II bindings on this locus, supporting the close interaction between PAF1 and RNA Poly II. A total of 52445 significant peaks (ChIP reads >= 20, background peaks < 20)

from CDC73 ChIP were mapped to the mouse genome. In order to characterize the distribution of binding sites, the genome-wide occupancy was separated into the promoter region (within 1kb up- or downstream of transcription start site), gene regions (including introns and exons) and intergenic regions. In order to define the overlap between LDB1, LMO2, CDC73 and POL-II, the relevant ChIP-Seq datasets ([19, 20] and unpublished data) were combined and genome-wide occupancy were compared. As shown in **Figure 2-a**, around 20% of CDC73, LMO2 and TIF1-γ, 35% of POL-II binding sites are localized at gene promoters, whereas LDB1 binding at the promoter region is only 5%. The binding sites of CDC73 within gene regions are quite similar to POL-II, again confirming the link between CDC73 and the PAF1 complex and basic transcription machinery. Based on the binding sites (peaks), the nearest target genes were identified (false discovery rate (fdr) < 0.001) for each transcription factor and compared with CDC73 co-occupancy (**Figure 2-b**). 56% of POL-II target genes were also bound by CDC73, and similarly 56% of LMO2 target genes were also bound by CDC73. However, 86% of LDB1 target genes were not co-occupied with CDC73, indicating that the majority of LDB1 target genes are not regulated together with CDC73 in MEL cells. 43% of TIF1-γ target genes also bound by CDC73, similar to results obtained in K562 cells [18]. Therefore, CDC73 may regulate its target genes together with POL-II, LMO2 and TIF1-γ.

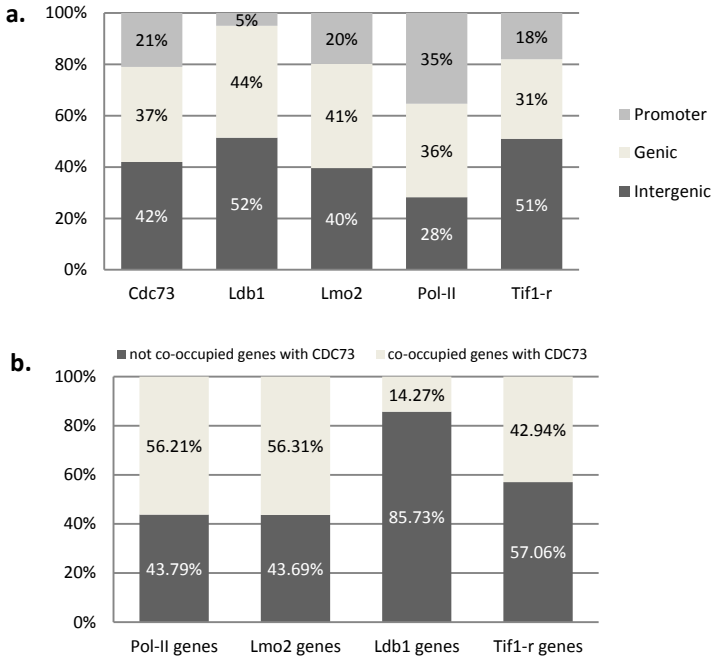


Figure 2. Genome-wide occupancy of transcription factors and target genes

a. Comparison of genome-wide binding sites of five transcription factors as indicated.

b. Co-occupancy of CDC73 target genes with other transcription factors

CDC73 together with LMO2 and LDB1 binds to erythroid specific genes

The number of CDC73, LMO2 and LDB1 target genes was 21191, 8135 and 3905 respectively. 14182 genes were uniquely bound by CDC73; 3915 genes and 290 genes were found common for LMO2 and LDB1, respectively. 2804 genes were found binding by all these three factors (**Figure 3**). Gene ontology analysis was performed on unique CDC73 target genes (14182), CDC73/LMO2 target genes (3915) and CDC73/LMO2/LDB1 target genes (2804). The top-10 significant functions are shown in the **Table 2**. The functions annotation column shows the prediction of cellular functions based on the target genes; while p-values give the significance.

Target genes binding CDC73/LDB1/LMO2 showed a preference for functions related to hematological disease, such as quantity and differentiation of red blood cells. In contrast, CDC73/LMO2 or CDC73/LDB1 target genes showed more functions related to gene expression, proliferation of cells. We also confirmed previous results that CDC73

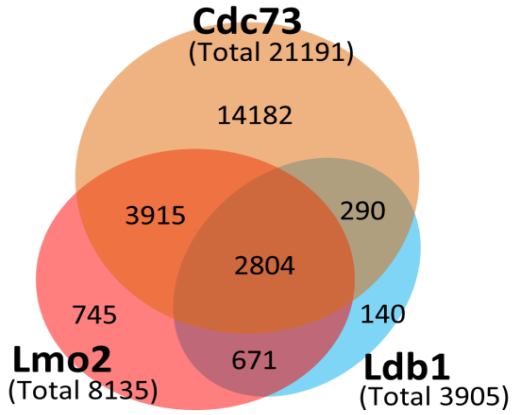


Figure 3. Venn diagram of CDC73, LMO2 and LDB1 target genes

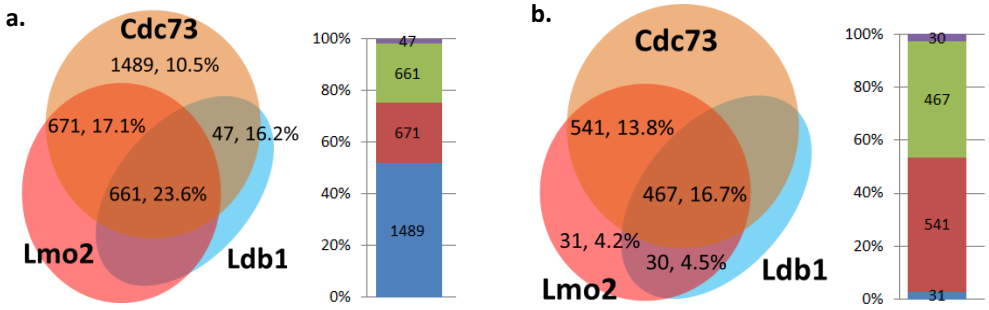


Figure 4. Venn diagram of CDC73 and LMO2 mis-regulated genes

Number of misregulated genes in CDC73 KDs (**a**) and in LMO2 KDs (**b**). Numbers are the misregulated genes, percentage indicates the ratio of the misregulated genes in each subgroups' target genes in Figure 3. The bar-graph indicates the percentage of each subgroups' misregulated genes in the sum of all subgroups.

Table 2. Gene ontology of common and unique CDC73, LDB1 and LMO2 target genes

	Functions annotation	p-Value
CDC73/LDB1/LMO2 (2804)		
1	Cell Death and Survival, cell death	3.73E-28
2	Cellular Growth and Proliferation, proliferation of cells	4.73E-26
3	Tissue Morphology, quantity of cells	2.46E-22
4	Hematological System Development and Function, quantity of blood cells	3.24E-22
5	Cellular Development, differentiation of cells	1.48E-21
6	Hematological System Development and Function, differentiation of blood cells	1.41E-19
7	Hematological Disease, anemia	1.60E-18
8	Infectious Disease, Viral Infection	6.09E-17
9	Cellular Function and Maintenance, cellular homeostasis	6.76E-17
10	Cellular Growth and Proliferation, proliferation of blood cells	4.07E-16
CDC73/LMO2 (3915)		
1	RNA Post-Transcriptional Modification, processing of RNA	1.48E-15
2	Cellular Growth and Proliferation, proliferation of cells	8.15E-12
3	Cell Cycle, cell cycle progression	3.09E-11
4	Gene Expression, expression of RNA	3.98E-11
5	Protein Synthesis, metabolism of protein	1.11E-09
6	Cell Death and Survival, cell death	2.54E-08
7	Developmental Disorder, Growth Failure	1.16E-06
8	DNA Replication, Recombination, and Repair, repair of DNA	2.26E-06
9	Cancer, lymphohematopoietic cancer	8.24E-06
10	Infectious Disease, Viral Infection	1.39E-05
Cdc73/Ldb1 (290)		
1	Cellular Growth and Proliferation, proliferation of cells	4.73E-06
2	Cell Cycle, meiosis II	1.11E-04
3	Cell Morphology, multinucleation of cells	1.20E-04
4	Cellular Compromise, multinucleation of cells	1.20E-04
5	Molecular Transport, quantity of lactic acid	1.50E-04
6	Infectious Disease, entrance of flavivirus	3.30E-04
7	Cellular Assembly and Organization, formation of microfibrils	3.30E-04
8	Tissue Development, formation of microfibrils	3.30E-04
9	Small Molecule Biochemistry, biosynthesis of 5,6,7,8-tetrahydrobiopterin	6.56E-04
10	Cancer, size of tumor cells	6.56E-04
CDC73 unique (14182)		
1	Gene Expression, transcription of DNA	4.86E-17
2	Organismal Survival, organismal death	2.80E-14
3	Embryonic Development, development of brain	5.07E-14
4	Behavior, behavior	1.14E-12
5	Embryonic Development, development of body axis	3.05E-10
6	Digestive System Development and Function, development of digestive system	6.87E-10
7	Embryonic Development, morphology of bone	6.69E-09
8	Embryonic Development, development of sensory organ	8.43E-09
9	Developmental Disorder, congenital anomaly of musculoskeletal system	1.03E-08
10	Cellular Assembly and Organization, disassembly of microtubules	1.53E-08
LMO2 unique (745)		
1	Inflammatory Response, inflammatory response	4.78E-07
2	Cellular Development, proliferation of lymphocytes	8.11E-07
3	Cellular Development, development of hematopoietic progenitor cells	1.01E-06
4	Cellular Function and Maintenance, flux of ion	2.14E-06
5	Cellular Development, proliferation of blood cells	2.23E-06
6	Cell-To-Cell Signaling and Interaction, activation of lymphocytes	2.63E-06
7	Hematopoiesis, maturation of leukocytes	2.86E-06
8	Connective Tissue Disorders, lupus erythematosus	5.99E-06
9	Hematological System Development and Function, emigration of phagocytes	9.45E-06
10	Hematological System Development and Function, activation of blood cells	1.01E-05

*The top-10 functions annotation for different target-gene groups were selected based on significance ($P \leq 0.05$).

binding sites are found in the proximity of cell-cycle related genes such as *Myb*, *Cyclin-D*, *Igf1r* genes or *c-Myc* and *Igf1* genes together with POL-II [21, 22]. Correspondingly, in the CDC73 unique target genes, we found developmental functions of diverse systems, for instance digestive system, development of body axis and sensory organ and morphology of bone. Thus CDC73 bind to loci with very different functions. This agrees with the observation that homozygous depletion of the *Cdc73* gene is embryonic lethal at E6.5 while a conditional deletion in adult leads to severe cachexia and death [23]. Therefore, our results also suggest that CDC73 may regulate genes that are important for embryonic development.

In conclusion, our ChIP results suggest that CDC73 as part of the PAF1 complex is the link between the LMO2 and LDB1 proteins to regulate hematopoietic and red-blood-cell related genes.

Knockdown of CDC73 or LMO2 shows transcriptional function as a repressor and an activator depending on their target genes

In order to identify genes whose expression depends on CDC73 and LMO2, we performed CDC73 or LMO2 knockdown with five shRNAs for each protein in MEL cells. The best knockdown levels of two shRNAs for each protein were selected (data not shown) and the RNA samples were prepared after three days of the knockdown of CDC73 or LMO2 and a scrambled shRNA as a negative control. The knockdown efficiency was -2.98 in log2-fold change (12.6% *Lmo2* RNA remained compared to scrambled shRNA) for LMO2 and -2.18 (22.1% *Cdc73* RNA remained) for CDC73. The resulting changes in transcriptome were analyzed using RNA sequencing (RNA-Seq). Combining the RNA-Seq data and ChIP-Seq data based on the binding sites, we found a number of genes that are specifically regulated by CDC73 (**Figure 4-a**) or by LMO2 (**Figure 4-b**), with a significance of log2-fold change range which is either higher than 0.6 or lower than -0.6. 1489 unique misregulated genes were found in CDC73 KD, comparing to the 14182 total target genes, it is only 10.5% genes have been changed. This may refer that CDC73 is not essential for the majority of its unique target genes, though the knockdown did not completely deplete CDC73. As shown in bar-graphs in **Figure 4**, it is clear that CDC73 KD showed more than 50% of misregulated genes (661 and 671) are also LMO2 and/or LDB1 target genes, which indicates that only half of CDC73 misregulated genes are regulated by CDC73/LMO2/LDB1 and CDC73/LMO2 complexes. The LMO2 KD showed the majority of misregulated genes (467 and 541) are regulated by LMO2/LDB1/CDC73 and LMO2/CDC73 complexes, respectively. These results suggest that in MEL cells, CDC73 together with LMO2 and/or LDB1 to regulate half of its target genes indicating the function of CDC73 in other complexes. Furthermore, the majority of LMO2 target genes are regulated by LMO2/CDC73 and LMO2/CDC73/LDB1 complexes.

The gene ontology of the misregulated genes in the CDC73 KD and LMO2 KD is shown in **Table 3**. We found that CDC73/LDB1/LMO2 misregulated genes in CDC73 KD (661 genes) are highly related to hematological functions such as quantity and differentiation of blood cells, but also contains the survival of organism function (**Table 3-a**), which genes in LMO2 KD showed the similar functions (**Table 3-b**). Among the misregulated genes of CDC73/LMO2 targets (671 in CDC73 KD; 541 in LMO2 KD), gene ontology analysis showed significant functions on cell death and proliferation related roles.

In order to find out the commonly misregulated genes in CDC73/LMO2 or CDC73/LMO2/LDB1, we compared the gene lists of 661 with 467, and 671 with 541 as shown in **Table S2**. We found that among 661 misregulated genes in CDC73KD in the CDC73, LMO2 and LDB1 targets, 202 genes are also misregulated in the LMO2KD. Among 671 misregulated genes in CDC73KD in the CDC73 and LMO2 targets, 220 genes are also misregulated in the LMO2KD. The gene symbols and log2-fold change are listed for 202 genes (**Table S2-a**) and 220 genes (**Table S2-b**). From the table, it is clear that the majority of these 202 and 220 genes are consistently misregulated in the same direction. If a gene such as *Alas2* goes up in LMO2 KD (1.093), it also goes up in CDC73 KD (1.763) and vice versa. Only a few genes showed a distinct regulation pattern such as *Cep72* and *Eprs* (**Table S2-b**). We further grouped these misregulated genes into consistently up- or down-regulated genes and performed gene ontology analysis as shown in **Table 4**. It showed that in both up- and down-regulated genes, their functions are highly related to cell death and proliferation. It is interesting to note that in the up-regulated genes, the microtubule dynamics is significant, which indicates that CDC73/LMO2/LDB1 may regulate the genes essential for microtubule organization involved in terminal erythroid differentiation [24]. In the down-regulated genes, we found the homeostasis of lipid function which is important during erythropoiesis [25]. We also found the significant function on abnormal morphology of proerythroblasts which indicates that CDC73/LMO2/LDB1 complex may maintain the erythroid cells in the premature stage.

In conclusion, CDC73 together with LMO2 and LDB1 regulates the genes for cell survival and proliferation in erythroid cells and also activates proerythroblast genes in the premature stages.

Table 3. Gene ontology for misregulated genes in CDC73 and LMO2 KDs

a. Misregulated genes in CDC73 KD

	Functions annotation	p-Value
CCD73/LDB1/LMO2 (661)		
1	Cell Death and Survival, cell death	3.09E-15
2	Cellular Development, differentiation of cells	5.39E-12
3	Tissue Morphology, quantity of cells	6.10E-11
4	Cellular Movement, leukocyte migration	1.75E-10
5	Cancer, Cancer	2.87E-10
6	Tissue Morphology, quantity of blood cells	3.59E-10
7	Inflammatory Response, inflammation of organ	9.22E-10
8	Cellular Development, differentiation of blood cells	4.37E-09
9	Cellular Movement, cell movement	5.63E-09
10	Cell-To-Cell Signaling and Interaction, adhesion of blood cells	1.29E-08
CDC73/LMO2 (671)		
1	Cell Death and Survival, cell death	2.22E-06
2	DNA Replication, Recombination, and Repair, DNA damage	9.52E-06
3	Connective Tissue Disorders, exencephaly	3.21E-05
4	Cell Morphology, repair of cells	4.45E-05
5	Free Radical Scavenging, removal of superoxide	5.19E-05
6	Cellular Growth and Proliferation, proliferation of cells	7.51E-05
7	Cancer, breast cancer	1.10E-04
8	Developmental Disorder, Growth Failure	1.23E-04
9	Hematopoiesis, quantity of hematopoietic cells	1.80E-04
10	Small Molecule Biochemistry, metabolism of oxalacetic acid	2.28E-04
CDC73 unique (1489)		
1	Amino Acid Metabolism, metabolism of amino acids	6.25E-08
2	Developmental Disorder, inborn error of amino acid metabolism	1.68E-05
3	Cellular Assembly and Organization, disruption of actin stress fibers	4.45E-04
4	Cell-To-Cell Signaling and Interaction, activation of hippocampal cells	9.11E-04
5	Metabolic Disease, acidemia	1.45E-03
6	Organismal Functions, thermogenesis	1.85E-03
7	Developmental Disorder, congenital generalized lipodystrophy	2.17E-03
8	Cancer, development of melanoma	2.30E-03
9	Cell Morphology, depolarization of cells	2.87E-03
10	Cancer, cytostasis of tumor cells	3.02E-03

b. Misregulated genes in LMO2 KD

	Functions annotation	p-Value
CDC73/LMO2/LDB1 (467)		
1	Cell Death and Survival, cell death	4,81E-12
2	Cellular Growth and Proliferation, proliferation of cells	8,08E-08
3	Cell Morphology, morphology of blood cells	6,82E-07
4	Hematological System Development and Function, morphology of red blood cells	7,10E-07
5	Cancer, Cancer	1,10E-06
6	Post-Translational Modification, phosphorylation of protein	1,33E-06
7	Hematological System Development and Function, quantity of blood cells	3,47E-06
8	Hematological Disease, anemia	3,90E-06
9	Molecular Transport, transport of neutral amino acid	6,26E-06
10	Cardiovascular Disease, occlusion of blood vessel	6,74E-06
CDC73/LMO2 (541)		
1	Cell Death and Survival, cell death	9,53E-07
2	Cellular Growth and Proliferation, proliferation of cells	1,68E-06
3	Carbohydrate Metabolism, metabolism of UDP-D-glucose	1,69E-05

4	Cellular Response to Therapeutics, sensitivity of cells	2,67E-05
5	Hematological System Development and Function, quantity of lymphocytes	4,08E-05
6	Cancer, breast cancer	6,92E-05
7	Embryonic Development, patterning of embryonic tissue	8,04E-05
8	Developmental Disorder, inborn error of carbohydrate metabolism	9,48E-05
9	Molecular Transport, localization of protein	1,48E-04
10	Cell Cycle, delay in interphase of cervical cancer cell lines	1,63E-04
LMO2 unique (31)		
1	Cellular Growth and Proliferation, colony formation of squamous cell carcinoma cell lines	8,40E-08
2	Immune Cell Trafficking, emigration of dendritic cells	2,01E-07
3	Developmental Disorder, autosomal recessive polycystic kidney disease	4,23E-07
4	Cellular Growth and Proliferation, proliferation of bone marrow cells	1,46E-06
5	Cardiovascular Disease, vasculitis	1,46E-06
6	Cancer, Joubert syndrome	1,94E-06
7	Hepatic System Disease, damage of liver	3,02E-06
8	Cellular Growth and Proliferation, proliferation of B lymphocytes	4,39E-06
9	Hematopoiesis, maturation of monocyte-derived dendritic cells	4,78E-06
10	Hematological System Development and Function, phagocytosis by macrophages	6,66E-06

*The top-10 functions annotation for different target-gene groups were selected based on significance (P ≤ 0.05).

Table 4. Gene ontology for commonly misregulated genes

	Functions Annotation	**p-Value**
Up-regulated genes in CDC73/LMO2/LDB1 and CDC73/LMO2		
1	Organismal Survival, organismal death	8,19E-08
2	Immunological Disease, allergy	1,58E-06
3	Hematological Disease, anemia	5,07E-06
4	Cellular Assembly and Organization, microtubule dynamics	9,50E-06
5	Free Radical Scavenging, quantity of reactive oxygen species	2,03E-05
6	Organ Morphology, contractility of cardiac muscle	2,23E-05
7	Hematological Disease, abnormal function of hematopoietic system	2,30E-05
8	Tissue Morphology, quantity of cells	2,42E-05
9	Immunological Disease, immediate hypersensitivity	3,84E-05
10	Inflammatory Response, inflammation of organ	5,51E-05
Down-regulated genes in CDC73/LMO2/LDB1 and CDC73/LMO2		
1	Cancer, autophagy of cancer cells	3,08E-05
2	Cell Morphology, abnormal morphology of red blood cells	3,80E-05
3	Hematological System Development and Function, quantity of blood cells	1,42E-04
4	Lipid Metabolism, homeostasis of lipid	1,62E-04
5	Cancer, tumorigenesis of malignant tumor	1,89E-04
6	Cell Morphology, abnormal morphology of proerythroblasts	3,54E-04
7	Cell Death and Survival, cell death	7,21E-04
8	Small Molecule Biochemistry, accumulation of alpha-amino acid	1,41E-03
9	Post-Translational Modification, activation of Protein kinase	1,73E-03
10	Free Radical Scavenging, removal of superoxide	1,87E-03

*The top-10 functions annotation for different misregulated-gene groups were selected based on significance (P ≤ 0.05).

Discussion

Here we determined the binding partners of LMO2 and found a novel partner CDC73, a component of the PAF1 complex component. ChIP-Seq and knockdown experiments of LMO2 and CDC73 showed that CDC73 forms a complex or complexes with the PAF1 and LDB1 complex to positively and negatively regulate target genes in erythroid cells.

In our results, we also found the Facilitates Chromatin Transcription (FACT) complex components in the LMO2 IP. It has been reported that the nucleosome remodeling FACT complex can be recruited by the PAF1 complex to promote transcription elongation [26]. Among the binding partners of LMO2, we also found TIF1-γ (**Table S1**), which recruits positive elongation P-TEFb complex to regulate the transcription of erythroid genes [18]. However, these factors are not found in the CDC73 IP MS. Interestingly, the HEXIM1 protein was found in CDC73 IP. It can bind to P-TEFb and inhibits the activity of P-TEFb and hence negatively regulate transcription elongation when binding with 7SK small nuclear RNA [27]. P-TEFb contains CDK9 which was studied previously in our lab [11]. In the CDK9 MS, the PAF1 complex components, PAF1, LEO1 and CDC73, the cell cycle related proteins AF4, CDK7, CyclinT1 and HEXIM1 are found. CDC73 was not only found in the LMO2-IP, LDB1-IP MS [1], but also in the GATA1-IP PAF1, CDC73 and LEO1 were found [12]. These results suggest that the PAF1 complex interacts with LDB1/LMO2/GATA1 complex.

We found CDC73/LDB1/LMO2 binding a number of important hematopoietic genes including the *Lmo2, Gata1*, *Fog1*(*Zfpm1*), *Tal1* and *Runx1* genes, and other erythroid related genes, such as *Alas2, Lyl1*, *Epb4.1* and *Klf2* (data not shown). The cell cycle related *Ccnd2*, *Ccnd3*, and Wnt signaling *Wnt3a* are also found in this subgroup, which suggests that these transcription factors also regulate cell growth, cell fate or play a role in oncogenesis. Among the misreguated genes of CDC73/LMO2, we found *Hexim1* which is downregulated in both the LMO2 and CDC73 KDs (**Table S2-b**). Thus the CDC73/LMO2 complex appears to activate *Hexim1* expression to negatively regulate elongation. These results indicate that CDC73/LMO2 complex may repress erythroid differentiation.

In CDC73 KD, we found a decrease of *Lmo2* gene expression (**Table S2-a**); while LMO2 KD did not influence *Cdc73* expression (data not shown). This indicates that *Cdc73* regulates *Lmo2* expression.

Myb gene is one of the cell-cycle related genes that is essential for cell proliferation and differentiation. The enhancer region of the gene has been previously studied by Stadhouders et al.[20], showing that *Myb* is one of the LDB1/LMO2 targets. However, we only find *Myb* in the category of CDC73 target genes. This is due to binding sites for

LDB1/LMO2 are closer to an unclassified cDNA sequence AK041716 (**Figure S3**). This cDNA is not expressed in the MEL cells [20]. We found CDC73 binding to the *Myb* promoter (**Figure S3**) and knockdown of CDC73 leads to activation of *Myb* (log2-fold change 1.18). In summary, the LDB1/LMO2 complex resides at the enhancer region of *Myb*, while the CDC73 protein occupies the promoter. We speculate that the combination of positive (LDB1/LMO2) and negative (CDC73) regulators determine the final *Myb* transcript levels.

Based on these results I propose the following model. LDB1/LMO2/GATA1 complex can form at least two complexes based on their target genes. One complex together with FACT, P-TEFb (CDK9 and CCNT1) and TIF1-γ to regulate genes that are essential for cell proliferation, while the other complex, containing CDC73, PAF1 complex and HEXMI1, activates genes maintaining erythroid cells at the proerythroblast stage. The choice of becoming a repressive or an active regulatory complex is probably dependent on one of the components of the LDB1/LMO2/GATA1 complex. We could not find any RNA polymerase pausing complexes, such as the DRB sensitivity inducing factor (DSIF) and the negative elongation factor (NELF), which exert the negative effect that has been replaced by NuRD complex on transcription elongation.

It should be noted that the target genes described in the **Figure3** and **4** are analyzed based on the binding sites. We did not discriminate the unique or different binding site on the same gene. As we found genes like *Gata2* that grouped in the common CDC73/LMO2/LDB1 target genes; however, the binding sites of these factors do not co-localize (data not shown). This suggests that CDC73, LMO2 and LDB1 may form distinct complexes that bind to different sites and regulate the same target genes probably via different mechanisms.

For further investigating the function of PAF1 complex on erythroid genes, it would be interesting to perform the ChIP-Seq for PAF1 complex components, for instance PAF1 and CTR9 to discover their target genes in MEL cells. It would be interesting to investigate the function of CDC73 upon erythroid differentiation, i.e. looking at the phenotype of MEL cell differentiation after *Cdc73* knockdown. The binding motifs on the up- and down-regulated genes would provide new binding partners that target the same specific genes. In summary, we reported here that CDC73 and PAF1 complex can interact with the LDB1/LMO2/GATA1 complex to activate genes essential for maintaining the premature erythroid stage.

Material and Methods

Cell lines and culture

Embryonic stem cells were cultured in the basic medium of DMEM (BE12-604F/U1, Lonza) containing 15% FCS, 100units/ml penicillin and 100mg/ml streptomycin (GIBCO/BRL), adding 1% non-essential amino acids (BE13-114E, Lonza), 0.00063% 2-mercaptoethanol (115433, Merck), 100U/ml Esgro (LIF) (ESG1106, Millipore). MEL cells were cultured in the basic medium.

RNA interference and gene expression profiling

RNA interference procedure was carried out followed a protocol which was previously described in details by Stadhouders [20]. shRNAs for Cdc73 and Lmo2 were obtained from SIGMA-ALDRICH MISSION shRNA library. shRNAs TRC number TRCN0000241652, TRCN0000241653 for Cdc73 and TRCN0000084313, TRCN0000084316 for Lmo2 were chosen for the RNA interference in RNA-Seq experiments. RNA was isolated from cells with either LMO2 or CDC73 knocked down at day 3 with the QIAGEN RNeasy Mini Kit, and integrity was checked on the Agilent 2100 Bioanalyzer. Then, the RNA sequencing was performed on Illumina HiSeq 2000 platform according to the manufacturer's instructions.

ChIP, ChIP sequencing and statistical analysis

Preparation of single cross-linking, sonication to 200-800-base-pair fragments, immunoprecipitations, and DNA purification was previously described in detail [12, 28]. Statistical analysis is described previously [19]. Primers for quantitative PCR are listed below.

Gata1 negative region:
Forward, TGCCGCTTGCCTTTGTAAG; Reverse, CACTAGCAGCTGGGTGGGTTA
Gata1 enhancer -3.5kb region:
Forward, TCAGGGAAGGATCCAAGGAA; Reverse, CCGGGTTGAAGCGTCTTCT
Gypa promoter:
Forward, CTCTGGTCCTCGCAGTTA; Reverse, CGTTGACACACATTGGCT
Alas2 promoter:
Forward, GGAACTGGGACATCTTGAC; Reverse, ACCATTAGAGTCTGGCTACT
Epb4.2 promoter:
Forward, CACTTGGCTTGAGTTCACAT; Reverse, GCTGCTGTGATGATTTCCC

Antibodies

The following antibodies were used in the Western Blot and ChIP experiments: anti-CDC73 (A300-170A, BETHYL). Anti-LDB1 (sc-11198), anti-TAL1 (sc-22809),

antibodies are from Santa Cruz and anti-LMO2 (ab72841) from Abcam.

Nuclear extract and Immunoprecipitation

Cell extract was prepared as described by de Boer et al. (de Boer et al., 2003) Briefly, cells were lysed in BufferA (10mM HEPES-KOH pH7.9, 1.5mM MgCl2, 10mM KCl, 0.5mM DTT, 1x Protease inhibitor) for ten minutes, followed by incubation in BufferB (20mM HEPES-KOH pH7.9, 25% Glycerol, 420mM NaCl, 1.5mM MgCl2, 0.2mM EDTA, 0.5mM DTT, 1x Protease inhibitor) for 30 minutes. The protein concentration was determined using Pierce® BCA Protein Assay (Thermo). 500µg total protein extract was used for each IP and incubated in Heng150 buffer (20mM HEPES-KOH pH7.9, 20% Glycerol, 0.25mM EDTA pH8, 0.05% NP40, 150mM KCl, 1x protease inhibitor) over-night at four degree. Protein A/G PLUS-Agarose beads (Santa Cruz Biotechnology, sc-2003) were blocked with CEA for one hour and added to IP samples for another hour of incubation. After washing with Heng150 buffer three times, 30µl Heng150 buffer was added to suspend the IP samples.

Western Blot

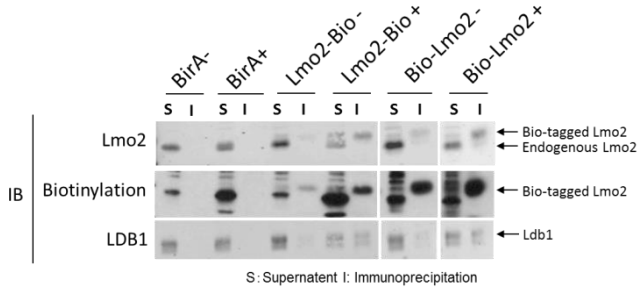
Samples in Laemmli buffer were boiled on a heat block for five minutes and loaded on NuPAGE precast 4-12% gradient Bis-Tris acrylamide gels (Novex). Proteins were transferred onto nitrocellulose transfer membrane (Whatman) and blocked with 1% milk in PBS-Tween. Detection of specific proteins was using the antibodies as specified. Fluorescent secondary antibodies were used for visualization by Odyssey system (LI-COR).

Mass Spectrometry

LMO2 and CDC73 IP samples in MEL cells were digested and dimethyl labeled as described by de Boer [29]. Samples were processed and analyzed by mass spectrometry (LS-MSMS, MS) as described previously [1, 12].

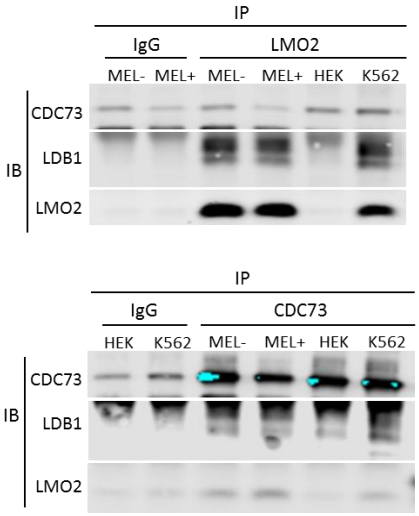
Supplementary data

Figure S1. LMO2 biotinylation and streptavidin pulldown.



LMO2, biotinylation and LDB1 detection on wester blot for LMO2-Bio and Bio-LMO2 clones under the non-induced (-) and induced (+) conditions in MEL cells.

Figure S2. Confirmation of CDC73 and LMO2 interaction in MEL cells



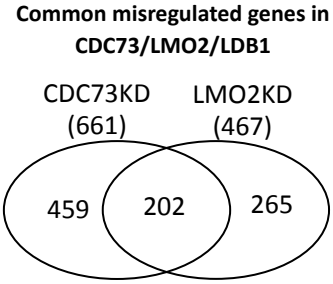
LMO2-IP (upper panel) and CDC73-IP (lower panel) in non-induced (-) and induced (+) MEL cells, HEK and K562 cells. All IPs were performed in 1 mg total nuclear extracts. CDC73, LDB1 and LMO2 proteins were detected.

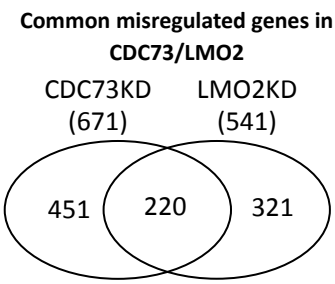
Table S1. Protein list of binding partners of LMO2 and CDC73 in MEL cells

*numbers indicate the mascot score

	LMO2 MS	CDC73 MS
LDB1 complex	LDB1(337), GATA1(106), LMO2(139), TAL1(667), TCF3(566), TCF4(343), TCF12(665), CBFA2T3(417), SSBP2(251), SSBP3(262), SSBP4(232), CBFA2T2(203), RUNX1(110), CDK9(49), LYL1(170), TRIM33(255), MYEF2(65)	LDB1(169), GATA1(54), LMO2(75), TAL1(200), TCF3(281), TCF12(269), CBFA2T3(285), SSBP2(81), SSBP3(114), SSBP4(127), CDK9(350)
	CDC73(77)	CDC73(2032), CTR9(3788), PAF1(2273), EO1(1895), WDR61(672)
NuRD complex	HDAC1(78), MTA2(46), LSD1(44)	HDAC2(55)
FACT complex	SUPT16H(281), SSRP1(156)	-
Polyadenylation complex	CPSF1(97)	CPSF2(882), CPSF3(627), CPSF4(248), CSTF1(76), CSTF3(353), SYMPK(181)
PTW/PP1 complex	PPP1CB(102), PPP1CA(153)	TOX4(603), PPP1R10(1007), PPP1CC(391), PPP1CB(357), PPP1CA(539), WDR82(573)
Others proteins	GFI1-B(68), EIF4A1(69), EIF4A3(85), EIF4ENIF1(191), FOP(128), IRF2BP2(129), NVL(131), POLR1A(46), HSPA1L(41), FXR1(114), RBMX(112), ZC3H14(45), DIDO1(119), PHF3(113), TRRAP(195), RBM15(209), RBBP6(168), FIP1L1(172), ZC3H11A(176), WDR11(44), MEF2(65), PSPC1(65), PRC1(185), RACGAP1(84), NONO(483), PRRC2C(135), RBM10(195), HSPA8(135), SNAP23(178), GPATCH8(93), PRRC2C(135), SMARCA5(144), CDC5L(131.5), DNAJC13(45), SF3B2(156), TOP2B(256), THOC5(58), RBM25(52), PGAM5(76), SUPT16H(281), SLTM(55), ILF3(68), THRAP3(507), UBAP2L(233), D1PAS1(65), KHDRBS1(71), SERBP1(79), SRBD1(91), NOM1(121), GM9000(45), ZFP281(154), TADA1(73), ACACB(1057), HIST1H1T(47), RPS6-PS1(87), GM8973(51), TOPBP1(113), D2WSU81E(68), RAN(66), 1700009N14RIK(66), ARPC4(58), KARS(53), YTHDF2(49), MNAT1(47), FTSJ3(45), MRE11A(44), GNL2(44), PLD1(42), PLEKHA7(44), NF1(44), PRDM2(52), HSPA2(41), HNRNPL(84), ZC3H4(48), IK(47), MMTAG2(80), CHCHD3(52), CHERP(81), IMMT(114), SNRPN(68), CHTF8(84), ATXN2L(84), THOC6(52), HSP90AB1(63), GM6472(69), PRRC2A(99), TRIM28(131), CSDA(250), YBX1(468), RUVBL1(73), DDX39B(81), FXR2(59), GM5292(187), GM5879(124), POLR1E(57), GM7263(77), NUFIP2(41), ABCF1(162)	HMGB1(261), HMGB2(171), HMGB3(74), HEXIM1(65), EIF4A3(275), FOP(70), CBFB(97), TCEB3(361), POLR1C(66), GTF3C1(72), ZBTB44(397), RIF1(261), CSNK2A1(198), RBBP6(487), PLEC(103), INTS1(181), NAA50(165), PPRC1(70), AZI1(901), MCG_118515(236), PRDX1(129), PCMT1(109), CLTC(139), MDH2(124), SRSF5(134), RCC1(146), RCC2(93), LYAR(199), THOC5(84), LUZP1(51), DEK(155), TARDBP(93), SAP18(62), GM10094(62), SMARCD1(78), TBL1XR1(58), G3BP1(254), RPL13A-PS1(89), C1QBP(83), GM6104(74), 2500003M10RIK(70), EIF3H(68), WTAP(64), INTS10(62), EXOSC5(62), PPIB(59), ALDH2(54), TRIM23(50), RBM25(547), ACIN1(436), RBM5(82), LUC7L3(161), CDK11B(247), PUF60(87), PPP2CB(60), ARF6(50), API5(98), SERBP1(190), GM20521(81), EIF5A(62), WDR33(898), WDR5(41), CUL5(111), PABPC6(116), CRBN(138), ZC3H13(94), DSP(61), PDIA3(94), ENO1(63), ARGLU1(51), CCT3(69), GTL3(114), INTS7(74), SLC25A3(50), INTS7(74), SLC25A3(50)

Table S2. Common misregulated gene list and gene ontology on these genes





a. 202 common misregulated genes in CDC73/LMO2/LDB1

	log2-fold change	
	CDC73	LMO2
Symbol	**KD**	**KD**
1700025G0		
4Rik	-1.164	-1.017
4632428N0		
5Rik	2.509	2.214
A930005H1		
0Rik	-1.267	-0.664
Aars2	0.622	0.609
Abcb10	1.054	1.307
Ache	0.814	0.753
Adcy6	0.752	0.893
Adcy7	0.664	0.797
Adcy9	-0.873	-0.842
Add3	-1.207	-1.502
Agrn	0.710	0.726
Akap7	-0.756	-0.869
Alas2	1.763	1.093
Ankrd27	-0.679	-0.657
Arhgap39	0.753	0.836
Arhgef10l	1.350	1.116
Atp1a1	1.330	0.692
B9d2	0.765	0.625
Banp	-0.732	-0.741
Bre	-1.175	-0.666
Btg2	1.084	1.825
Calm2	1.149	0.721
Camta1	1.300	1.294
Cbr4	-1.123	-1.252
Ccdc112	-0.824	-0.947
Ccdc130	-0.993	-1.364
Cd53	-1.156	-1.365
Cdk2ap2	0.786	0.672
Cdkn1a	1.092	0.961
Cdkn2b	-1.046	-0.818
Cenpa	0.783	0.866
Cep55	-1.350	-0.768
Coro2a	-1.119	-0.919
Cpox	1.181	0.866
Creb3l1	-1.041	-0.999
Cryzl1	-1.302	-1.334
Cyth4	0.653	0.652
D14Abb1e	-0.953	-0.799
Ddit3	-1.600	-2.646
Ddost	0.827	1.040
Def8	1.088	1.220
Dnaja3	-1.537	-0.777
Dyrk3	1.101	1.507
Ecm1	1.284	1.280
Ehd1	0.749	0.763
Emilin2	1.118	1.279
Epn2	-0.705	-0.729
Fads2	1.486	2.013
Fam102a	0.784	0.773
Fam46c	0.998	0.785
Fen1	0.695	0.618
Fignl1	-0.707	-0.842
Fut8	-0.648	-0.804
Galk1	0.632	0.908
Galnt10	0.842	0.791
Gata2	1.080	0.633
Gfpt2	0.743	0.962
Gng2	-1.062	-2.158
Gps2	-0.689	-0.635
Grk5	1.661	1.148
Gsg1l	-2.386	-0.632
Gsr	1.152	1.090
Gtpbp2	-1.846	-0.965
Hdac7	1.120	0.983
Hdgf	0.960	0.605
Hecw1	-1.213	-0.862
Hipk1	0.854	0.720
Hmgcr	0.752	0.927
Hmgcs2	1.427	2.104
Hnrnph3	-0.989	-0.894
Hsp90aa1	1.482	0.690
Hspa5	1.236	1.466
I830077J02		
Rik	-1.039	-1.718
Ifngr2	-0.644	-0.688
Inf2	0.936	1.123
Insig1	1.356	1.562
Itgb3	1.611	1.080
Itgb7	0.771	1.160
Itprip	0.770	0.734
Kcnh2	-1.701	-1.875
Kcnn4	1.013	0.652
Kif26a	0.656	0.605
Kifc3	0.947	1.171
Kit	1.554	0.991
Klf1	0.701	0.805
Krcc1	-1.458	-1.838
Ldlr	2.605	4.065
Ldlrap1	0.635	1.081
Lgals8	-0.951	-1.774
Lgals9	1.491	1.284
Lmna	0.922	1.024
Lmo2	-1.192	-2.986
Lrrfip1	-0.611	-0.626
Lyst	-0.873	-0.775
Map2k3	1.607	0.699
Mapkapk2	0.786	0.741
Mfsd6	-0.777	-0.914
Micall2	0.703	0.833
Mknk1	-1.112	-0.769
Mknk2	2.157	1.026
Mrm1	-0.841	-1.098
Mst1	-2.987	-1.053
Myl9	0.844	1.304
Myo6	-0.853	-1.583
N4bp2l1	-0.898	-0.628
Nars	-0.740	-1.175
Ncoa1	-0.629	-0.659
Nfam1	0.819	0.725
Nfe2l2	0.888	0.867
Nmnat3	-1.525	-1.623
Nos3	1.160	1.138
Nupr1	-1.414	-2.931
Ociad2	0.603	0.644
Odf2l	-1.015	-0.964
Olfm1	1.337	0.768
P2ry14	-1.199	-1.762
Pcyt1b	0.814	2.088
Pdcd1lg2	-0.701	-0.763
Pik3cg	0.669	1.003
Pkd2l2	-0.775	-1.348
Pkhd1l1	0.693	1.215
Pml	0.723	0.839
Pomp	0.920	0.871
Popdc2	1.820	0.606
Ppapdc1b	-0.989	-0.706

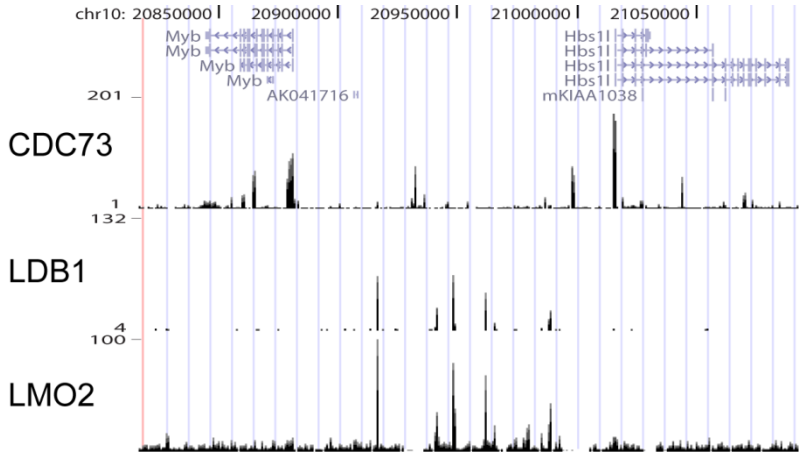
Prr5	0.676	1.305
Ptgs1	1.925	1.934
Ptprs	0.909	0.635
Pvrl2	0.856	1.733
Rbm45	-0.703	-1.414
Rchy1	-0.859	-0.861
Rec8	1.425	1.942
Relb	-2.154	-2.525
Rin3	1.057	0.898
Rnf145	0.705	0.724
Rwdd3	-0.661	-0.954
Ryk	0.629	0.934
Scarb1	0.671	0.897
Scd2	1.014	1.676
Sec61a1	0.821	0.849
She	-1.681	-1.051
Shox2	0.705	1.107
Sik1	1.086	1.093
Sipa1	0.920	1.093
Sipa1l1	0.633	0.864
Slc14a1	0.799	0.821
Slc22a23	0.868	0.611
Slc22a3	2.005	1.242
Slc26a1	2.622	0.910
Slc30a10	0.946	0.954
Slc43a2	0.710	0.889
Slc45a4	1.111	0.683
Slc6a6	1.102	1.188
Slc6a9	-1.430	-1.271
Slc7a1	-0.967	-1.468
Smox	-1.012	-1.269
Soat2	-1.982	-1.713
Sos1	-0.714	-1.028
Spata13	0.657	1.015
Spg11	-0.710	-1.037
Spns2	1.013	2.760
Sqle	0.892	0.975
St3gal6	-0.727	-0.765
St6galnac3	-2.163	-0.961
Steap3	0.939	0.818
Tbc1d5	-1.839	-1.443
Tbck	-0.744	-0.728
Tmco4	-1.173	-0.669
Tmem161b	-0.617	-0.781
Tmem194b	-0.609	-0.942
Tnni3	1.740	1.629
Tnrc6b	-1.168	-0.611
Tor3a	-1.055	-1.264
Trib3	-1.188	-2.953
Tspo2	-1.652	-0.974
Tspyl3	1.287	0.763
Ttyh3	0.852	0.969
Tuba4a	1.351	0.634
Ugcg	1.549	0.698
Vangl1	0.848	0.613
Vcl	1.200	0.783
Wdr67	-1.117	-1.143
Wdr7	-0.651	-0.716
Zcchc7	-0.867	-0.947
Zdhhc24	-0.710	-0.605
Zfp740	-0.827	-0.773
Zfr	-0.819	-0.654
Zyx	0.656	0.787
Zzz3	-0.696	-0.748

b. 220 common misregulated genes in CDC73/LMO2

	log2-fold change	
	CDC73	LMO2
Symbol	**KD**	**KD**
1500012F01Rik	-1.429	-1.161
1810032O08Rik	1.512	0.906
2010204K13Rik	-1.387	-1.033
2210404J11Rik	-0.952	-1.004
2410006H16Rik	-2.117	-1.561
4930404N11Rik	0.642	1.144
4930432K21Rik	0.676	1.817
4932415G12Rik	-2.112	-2.878
5430416N02Rik	-1.711	-1.812
9430008C03Rik	-1.858	-1.769
A630001G21Rik	0.736	0.675
Aacs	-0.739	0.605
Aasdh	-1.248	-0.804
Abhd11	-1.022	-1.049
Acat2	-0.605	-1.072
Acly	1.371	1.016
Adal	-0.695	-1.350
Adam15	-0.697	-0.657
Aggf1	0.613	0.997
AI413582	-1.360	-0.998
Aldh2	0.946	1.208
Amdhd2	0.868	0.926
Angptl6	-1.009	-0.680
Ank2	-0.821	-1.179
Anks3	1.107	0.981
Arhgap12	-1.030	-0.917
Asns	-0.748	-1.116
Atf5	-0.997	-1.354
Atp1b2	0.821	0.755
Atp2b4	1.037	0.819
Atp7a	-0.838	-0.609
Atxn7	1.253	0.634
AW549877	-1.036	-0.625
BC005537	0.647	0.903
BC055111	0.714	0.606
Bcdin3d	0.850	1.137
C030006K11Rik	-1.536	-1.836
C330006A16Rik	-2.264	-2.587
C920021L13Rik	-1.678	-0.829
Calm3	1.728	1.003
Camkk1	0.632	1.335
Capn1	1.054	0.864
Caprin2	-0.756	-0.997
Casp3	-0.786	1.244
Cd24a	-1.122	-0.672
Cebpg	-1.256	-0.850
Cep72	0.615	-0.836
Cetn3	0.616	0.623
Chd2	-1.508	-1.484
Chd5	0.685	0.605
Clstn3	1.921	0.888
Coro1a	0.920	0.962
Cox7a2	-0.811	0.815
Crip1	2.210	1.251
Ctbs	-0.709	-0.692
Cyb5d2	-1.294	-1.223
Cyb5r1	-0.654	-0.626
Ddx11	-0.914	-1.093
Dgcr6	0.686	0.814
Dhcr24	-0.881	-1.313
Dhcr7	-1.298	-1.217
Dhx38	-0.997	-1.146
Dip2b	-0.603	-0.981
Dnajb1	0.683	0.758
Dock2	0.642	0.769
Dok2	1.240	1.404
Dpy30	-0.614	-0.708
Dscr3	0.600	0.724
Dus2l	-0.791	-0.705
Dvl2	1.472	0.684
Dync2h1	-1.123	-1.088
Efhd2	-1.029	-1.017
Egln2	-1.339	-0.664
Endog	0.674	-0.894
Entpd7	-0.637	-1.923
Epc1	-0.621	-0.688
Eprs	-0.686	0.708
Fam118b	-1.159	-1.314
Fam129a	-0.659	-0.608
Fam175a	0.872	0.850
Fancc	1.284	0.858
Fasn	-0.627	-0.749
Fbxl19	0.914	0.869
Fdxr	0.657	1.124
Fis1	0.654	0.667
Fkbp4	0.923	0.867
Flna	-0.785	-1.008
Foxh1	-0.841	-0.984
Gclm	0.987	1.085
Glul	-0.655	-0.664
Gm10785	-1.733	-1.489

Gm11974	-1.718	-2.334
Gnai2	-0.819	-0.855
Gng10	0.846	0.740
Got1	-0.630	-1.224
Gpr19	0.888	1.350
Gpt2	0.900	1.033
H2afx	-1.155	-0.910
H2-DMa	-0.984	-1.220
H3f3b	0.674	-0.601
Hax1	0.988	0.865
Hemk1	0.888	1.129
Hexim1	-1.138	-1.048
Hjurp	0.752	0.611
Hnrnpab	0.638	0.655
Hspbap1	0.605	1.638
Hsph1	0.941	0.945
Id1	-1.038	-2.772
Ift172	0.847	0.967
Igfbp4	0.672	1.980
Irf1	0.654	0.984
Isyna1	0.645	0.622
Kdelr2	-1.614	-0.846
Kiss1r	-0.909	-1.003
Krtcap2	0.807	0.889
Lactb2	1.050	0.797
Letm2	-0.983	-1.317
Lif	0.880	-0.613
Lmo4	0.961	0.630
Lonp1	0.828	0.993
Lpcat2	0.883	0.650
Lrp8	-0.713	-0.875
Lsm7	-0.886	-0.693
Lss	-0.828	-0.627
Lyrm2	-1.110	-1.359
Lyrm5	-1.015	-1.288
Malat1	-2.958	-2.244
Maml1	-1.175	-1.689
Mat2a	0.732	0.748
Mfsd11	0.626	0.679
Mpdu1	1.896	2.158
Mrpl28	-0.620	0.613
Mthfd2	0.678	0.607
Mxd4	-0.975	-0.985
Mybl2	0.672	1.421
N6amt1	0.779	0.893
Nagk	-0.861	-0.657
Ncf2	1.034	1.352
Ndufa4	-0.746	-1.063
Ndufa4l2	-1.015	-0.737
Ndufb7	0.881	1.323
Neat1	-2.281	-2.113
Nin	0.622	0.610
Nlgn2	0.744	-0.780
Npm1	-1.271	-1.543
Nt5dc2	-1.423	-1.780
Ntpcr	-0.790	-1.040
Nudc	-0.727	-1.133
Ogfr	-1.144	-0.671
Osbpl1a	0.798	0.665
Osbpl2	0.837	1.068
Pak6	-1.433	-1.517
Palm	-0.934	-1.701
Paqr3	-1.242	-1.842
Paqr7	-0.980	-0.674
Parvg	1.178	1.280
Pcbd2	1.262	0.688
Pck2	-1.022	-1.340
Pcyt2	-0.626	-0.885
Pfn1	1.897	0.869
Pgp	-1.054	-1.611
Phlda2	0.675	0.848
Pim2	-0.782	-0.740
Prdx1	-0.721	-0.863
Prkdc	0.608	-0.656
Ptpru	-0.731	-1.508
Pts	0.896	1.035
Rad18	0.925	0.620
Rad50	0.635	1.209
Rad51	1.031	1.406
Rad54l	-0.725	-1.187
Rbm39	0.992	1.446
Rpl22l1	0.836	0.652
Rpp40	1.289	1.312
S1pr2	-1.108	-0.692
Sbk1	-1.053	1.148
Scfd1	1.343	0.835
Sema4a	-0.703	-0.807
Sgpl1	0.644	0.824
Shank1	0.841	0.776
Slc17a5	-1.153	-1.020
Slc1a4	1.649	0.928
Slc38a1	1.105	1.023
Slc48a1	-1.701	-1.856
Slc9a3r1	0.634	1.166
Smarca4	0.892	1.206
Snhg12	-2.011	-0.843
Spred1	-0.696	-1.343
Srd5a3	0.950	0.800
Srebf2	0.612	1.260
Stard10	-0.780	-0.632
Taldo1	-0.648	-0.627
Tatdn1	0.745	0.651
Tdrd9	0.727	1.098
Tecpr1	-1.403	-1.086
Tfrc	0.607	-0.646
Tmem106c	-1.188	-1.026
Tmem126a	-0.777	-0.728
Tmem209	0.932	0.841
Tomm40	0.688	1.763
Tprkb	-1.222	-0.814
Trim46	-1.041	-1.680
Ttll11	-0.691	-0.781
Tuba1b	-0.602	-1.082
Tulp4	-0.844	-0.826
Twf2	1.089	0.793
Usp40	2.292	1.881
Vav2	0.676	0.958
Vps8	-0.836	-2.254
Wbp1	-0.753	-1.299
Wdr6	-1.396	-0.759
Ypel5	0.833	0.709
Zbtb45	-1.151	-1.296
Zc3h8	0.996	0.791
Zcwpw1	-0.965	-0.676
Zdhhc21	0.954	0.660
Zfp420	-1.295	-0.717
Zfp667	0.708	0.689
Zfp839	1.295	0.978
Zmat1	-1.203	-1.240

Figure S3. ChIP-Seq of CDC73, LDB1 and LMO2 in the Myb-Hbs1l locus in MEL cells



References

1. Meier, N., et al., *Novel binding partners of Ldb1 are required for haematopoietic development.* Development, 2006. **133**(24): p. 4913-23.
2. Patterson, L.J., et al., *The transcription factors Scl and Lmo2 act together during development of the hemangioblast in zebrafish.* Blood, 2007. **109**(6): p. 2389-98.
3. Warren, A.J., et al., *The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development.* Cell, 1994. **78**(1): p. 45-57.
4. Boehm, T., et al., *The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13.* Proc Natl Acad Sci U S A, 1991. **88**(10): p. 4367-71.
5. Terano, T., et al., *Transcriptional control of fetal liver hematopoiesis: dominant negative effect of the overexpression of the LIM domain mutants of LMO2.* Exp Hematol, 2005. **33**(6): p. 641-51.
6. Visvader, J.E., et al., *The LIM-domain binding protein Ldb1 and its partner LMO2 act as negative regulators of erythroid differentiation.* Proc Natl Acad Sci U S A, 1997. **94**(25): p. 13707-12.
7. Hansson, A., et al., *The Lim-only protein LMO2 acts as a positive regulator of erythroid differentiation.* Biochem Biophys Res Commun, 2007. **364**(3): p. 675-81.
8. Chao, S.H. and D.H. Price, *Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo.* J Biol Chem, 2001. **276**(34): p. 31793-9.
9. Peng, J., et al., *Identification of multiple cyclin subunits of human P-TEFb.* Genes Dev, 1998. **12**(5): p. 755-62.
10. Ni, Z., et al., *P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo.* Mol Cell Biol, 2008. **28**(3): p. 1161-70.

11. Bezstarosti, K., et al., *Differential proteomics based on 18O labeling to determine the cyclin dependent kinase 9 interactome.* J Proteome Res, 2010. **9**(9): p. 4464-75.
12. Rodriguez, P., et al., *GATA-1 forms distinct activating and repressive complexes in erythroid cells.* EMBO J, 2005. **24**(13): p. 2354-66.
13. Wang, Y., et al., *SSBP2 is an in vivo tumor suppressor and regulator of LDB1 stability.* Oncogene, 2010. **29**(21): p. 3044-53.
14. Bai, X., et al., *TiF1-gamma plays an essential role in murine hematopoiesis and regulates transcriptional elongation of erythroid genes.* Dev Biol, 2013. **373**(2): p. 422-30.
15. van Riel, B., et al., *A novel complex, RUNX1-MYEF2, represses hematopoietic genes in erythroid cells.* Mol Cell Biol, 2012. **32**(19): p. 3814-22.
16. Snow, J.W. and S.H. Orkin, *Translational isoforms of FOG1 regulate GATA1-interacting complexes.* J Biol Chem, 2009. **284**(43): p. 29310-9.
17. Mueller, C.L. and J.A. Jaehning, *Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex.* Mol Cell Biol, 2002. **22**(7): p. 1971-80.
18. Bai, X., et al., *TIF1gamma controls erythroid cell fate by regulating transcription elongation.* Cell, 2010. **142**(1): p. 133-43.
19. Soler, E., et al., *The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation.* Genes Dev, 2010. **24**(3): p. 277-89.
20. Stadhouders, R., et al., *Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development.* EMBO J, 2012. **31**(4): p. 986-99.
21. Lin, L., et al., *The parafibromin tumor suppressor protein inhibits cell proliferation by repression of the c-myc proto-oncogene.* Proc Natl Acad Sci U S A, 2008. **105**(45): p. 17420-5.
22. Yang, Y.J., et al., *The tumor suppressor, parafibromin, mediates histone H3 K9 methylation for cyclin D1 repression.* Nucleic Acids Res, 2010. **38**(2): p. 382-90.
23. Wang, P., et al., *Parafibromin, a component of the human PAF complex, regulates growth factors and is required for embryonic development and survival in adult mice.* Mol Cell Biol, 2008. **28**(9): p. 2930-40.
24. Freikman, I., I. Ringel, and E. Fibach, *Shedding of phosphatidylserine from developing erythroid cells involves microtubule depolymerization and affects membrane lipid composition.* J Membr Biol, 2012. **245**(12): p. 779-87.
25. Kuypers, F.A., *Membrane lipid alterations in hemoglobinopathies.* Hematology Am Soc Hematol Educ Program, 2007: p. 68-73.
26. Formosa, T., *The role of FACT in making and breaking nucleosomes.* Biochim Biophys Acta, 2012. **1819**(3-4): p. 247-55.
27. Dey, A., S.H. Chao, and D.P. Lane, *HEXIM1 and the control of transcription elongation: from cancer and inflammation to AIDS and cardiac hypertrophy.* Cell Cycle, 2007. **6**(15): p. 1856-63.
28. Soler, E., et al., *A systems approach to analyze transcription factors in mammalian cells.* Methods, 2011. **53**(2): p. 151-62.
29. de Boer, E., et al., *Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice.* Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7480-5.

Chapter 4

Discussion

General discussion

The differentiation of hematopoietic stem cells (HSC) into the erythroid lineage involves extrinsic and intrinsic signals. Functionally conserved transcription factors such as *Gata2*, *Runx1*, *Ldb1*, *Tal1*, *Lmo2* and *Gata1* are essential for this differentiation [1]. GATA1 is a key regulator in the development and differentiation of erythroid [2], megakaryocytic [3], eosinophilic [4] and mast cell [3] dendritic cell lineages. It can bind a number of factors including FOG1, LDB1 and GFI-1b to form at least five distinct complexes [5].

GATA1/LDB1 complex and early hematopoiesis

The ubiquitously expressed nuclear protein LDB1 does not bind to DNA but functions as a scaffold in the GATA1/LDB1 complexes. It has been shown that LDB1 is involved in long-range interactions, for example the binding sites of the LCR (locus control region) are spread around a 100 kb from the active human or mouse *β-globin* gene and LDB1 has been shown to be important to connect these sites [6]. Similar long-range interactions mediated by the GATA1/LDB1 complex occur in the *α-globin* cluster [7]. The pull-down of LDB1 followed by mass spectrometry unveiled in addition to already known partners such as LMO2 and TAL1 new binding partners in the GATA1/LDB1 complex such as CDK9 and ETO2, which were shown to be critical for definitive hematopoiesis in zebrafish [8]. Immunohistochemical analysis showed that the GATA1/LDB1 complex components are co-expressed in the same cells in the para-aortic splanchnopleura (P-Sp) region on embryo sections at E9.5, a region that gives rise to definitive hematopoiesis in the AGM at later stages. It indicates that the GATA1/LDB1 complex is not only important in a late mature cell type, but is already present at the earliest stages of hematopoiesis [8]. However, it is still not known whether these factors are only present in the early stage as single protein or that they also function as a complex to regulate genes. In order to discover the temporal and spatial emergence of the GATA1/LDB1 complex during early embryonic development, we performed PLA experiments (Chapter 2) on embryoid bodies. This showed the GATA1/LDB1 complex to appear at day 4 of ES cell differentiation. By contrast, the GATA1/FOG1 complex emerged at day 5. This sequential emergence of distinct GATA1 complexes agrees with earlier observations that they form different complexes with different functions. It was previously suggested that the GATA1/FOG1 complex together with the MeCP1 complex represses alternative lineage genes in the early erythroid cells, while the GATA1/LDB1 complex including TAL1 would activate genes essential for later stages of erythroid differentiation. Our results in the *in vitro* ES cell differentiation system show that the complexes appear before the appearance of hemangioblasts suggesting that there are also early (unknown) functions of the GATA1/LDB1 complex in the early embryonic development before the emergence of

HSCs. The LDB1/E2A interaction already exists in the EBs at day 4. However due to the low amount of GATA1 protein in the EBs, we could not detect it in the LDB1 precipitation and hence we cannot conclude that LDB1/E2A and GATA1/LDB1 are already in the same complex.

The next question would be in which cell types the GATA1 complexes are formed, and what their functions are. It would be interesting to see whether the complete GATA1/LDB1 complex exists in the Flk1+ cell population in the EBs, which represents the endothelial-hematopoietic progenitors, BL-CFCs.

Ldb1-KO ES cell differentiation shows a lack of hematopoiesis in yolk-sac due to a decreased number of BL-CFCs [9]. Genome-wide analysis of LDB1 binding sites in Flk1+ cells revealed its target genes, including *Runx1*, *Gata2*, *Gata1*, *Eto2* and *Tal1*, which are essential for early hematopoietic and endothelial development, and which are downregulated in the Ldb1-/- Flk1+ cells [9].

Since the GATA1/LDB1 complex together with MeCP1 represses *Gata2* expression in erythropoiesis during GATA factor switching, we assume that either the LDB1 complex contains other proteins that activate the *Gata2* gene in Flk1+ cells, or that it may form a distinct complex to upregulate its expression. Motif analysis of Ldb1 peaks in Flk1+ cells from BL-CFCs and adult hematopoietic Lin- bone marrow progenitors reveals that the LDB1 binding motif is associated with the CTCF motif in Flk1+ cells, while the peaks in Lin^- cells show a prominent E-box:Gata motif and no CTCF motif [9]. The multimeric GATA1/TAL1/LDB1/E2A/LMO2 complex binds to the E-box:Gata motif to activate erythroid genes, such as *glycophorin A* [10] and *α-globin* [7] during erythropoiesis. These results suggest that LDB1 forms distinct complexes in the early stage of embryonic development compared to the mature erythroid stage, probably with GATA1 or together with E2A to regulate the early hematopoietic genes in the Flk1+ cells or that it uses different co-factors binding in close proximity. Proteomics analysis of LDB1 binding partners in the Flk1+ cells after benzonase treatment would provide complete interaction network to uncover the function of LDB1 complexes in the early embryonic development (currently performed by A. Martella in the lab), while an analysis without benzonase treatment may uncover co-factors binding to the DNA in proximity to the GATA1/LDB1 complex. Alternatively components of the complex may carry different post-transcriptional modifications, which could be uncovered using mass labeling followed by mass spectrometry.

Gata1 knockout mouse embryos show severe anemia and die at E10.5 to E11.5 due a blocking in proerythroblast differentiation, leading to the apoptosis [11, 12]. Paradoxically, overexpression of *Gata1* in erythroid cells also inhibits their differentiation and leads to a lethal anemia [13]. A comparative experiment of the potential of *Gata1, Gata2* and *Gata3* transgenes to rescue the Gata1-null mice showed that the quantitative levels of GATA1 are important for erythroid

differentiation [14]. They also suggest that *Gata1* target genes may respond to different threshold levels of GATA factors, i.e. globin genes may need high levels of GATA1, whereas genes regulating cell survival already respond to a low level [15]. This phenomenon may also exist in the early development of hematopoietic cells. Some early hematopoietic genes may be mediated by a low level of GATA1 in the complex of GATA1/LDB1. Once these genes are activated, they can push the stem cells differentiating into the hematopoietic committed stage, after which GATA1/LDB1 complex will further activate *Gata1* expression, forming a feed forward loop for further lineage specific development.

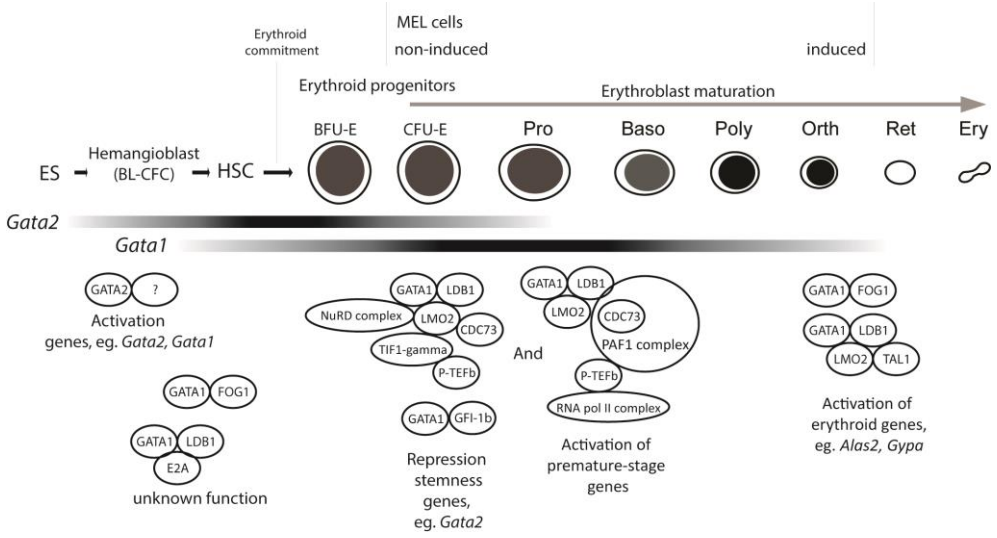
Transcription factor complexes and RNA polymerase II

Cell-cycle progression related CDK9 has been precipitated in the LDB1-IP only in the non-induced MEL cells [8] suggesting the link between LDB1 complex and CDK9 to maintain proliferative state. Chromatin remodeling complex components such as HDAC1 also showed a decreased interaction with LDB1 in induced MEL cells. These results suggest that the GATA1/LDB1 complex bringing in CDK9 maintains the proliferative state of erythroid cells. Perhaps this also explains that the *cdk9*-MO injection experiment resulted in a reduction of body size of zebrafish embryos [8]. A CDK9 IP in MEL cells found an interaction with the Mediator complex component MED1. This would agree with a recent report that MED1 facilitates GATA-1-dependent transcription to regulate erythropoiesis [16]. In Chapter-3, I describe a newly identified binding partner of LMO2, the CDC73 protein which is a component of PAF1 complex, which together with the LDB1 complex can repress differentiation genes and activate genes for the premature stage of the erythroid lineage.

Our results are consistent with the previous observations and provide an additional link between the transcription factor complex and the basal RNA polymerase II complex. The PAF1 complex is involved in a diverse and essential role: for instance, CDC73 and CTR9 are important for transcriptional elongation; PAF1 and RTF1 communicate with transcription factors and LEO1 and SKI8 are important for recruitment and activation of histone modification factors [17]. In the previously identified proteins involved in γ-globin silencing, the zinc finger protein ZBP-89 has found [18]. It was shown to interact with GATA1 and MAFK to regulate erythroid development [19]. It has also been reported that ZBP-89 interacts with HDAC1 to repress its target genes such as γ-globin [20]. In conclusion, I suggest that CDK9 and CDC73 proteins can link the LDB1 complex with the basal RNA Pol II complex via CDC73 and CDK9 to activate the early/premature stage of the erythroid lineage and repress differentiation genes.

I have summarized my current understanding of the role of different GATA1 transcription complexes during the hematopoiesis in the **Figure**. During the early embryonic development, GATA2 with as yet unknown partners activates the *Gata2* and *Gata1* genes. Once GATA1 is expressed, GATA1 will bind to LDB1 and later to FOG1 to regulate their target genes. After erythroid commitment, at the early erythroid stage, GATA1 forms three complexes: one complex is GATA1/LDB1/NuRD/TIF1-γ that represses the stemness genes such as *Gata2*. Another one is GATA1/GFI-1b to repress cell proliferation related genes such as *Myc* and *Myb*. The last complex is GATA1/LDB1/PAF1 to activate premature-stage genes to prevent differentiation. Erythroid differentiation is started by increasing levels of GATA1 forming two major complexes, GATA1/FOG1 and GATA1/LDB1/LMO2 complex to activate erythroid specific genes such as *Alas2* and *Gypa*.

Figure. Schematic presentation of distinct GATA1 complexes during the hematopoiesis



Optimization and perspective of PLA technique

Compared to other methods for protein-protein interaction measurements, for instance, co-IP followed by mass spectrometry, yeast two-hybrid, FRET/BRET and BiFC, *in situ* PLA provides the outstanding advantage to detect protein-protein interactions even when few cells can be obtained. Since it can detect up-to 40nm distance for a protein-protein interaction, its high resolution provides more detailed information when compared to normal confocal immunofluorescence which can only distinguish the distance of few-hundred nanometers.

In order to perform the *in situ* PLA experiments, we first tested different primary antibody dilutions in the immunofluorescence (IF) stainings with the same concentration of secondary antibodies. It should be noted that the fixation reagent should be optimized for the antibodies. Generally, 2-4 % Paraformaldehyde (PFA) fixation works for majority of the antibodies if samples are incubated in a short time. The permeabilization time should be also short especially for suspension cell samples. In my hands, after dissection, embryonic tissues are embedded in the OCT reagent and are frozen directly. The sample sectioning is performed later. It may be better to fix the samples first then go to the OCT-freezing step. The sectioned samples should not be kept too long in the -80 degree freezer. Based on the IF signals, we chose the most optimal concentration for the antibody, clear enough to see the staining but not over-stained. Otherwise, the ratio of signal versus noise will become low. (The noise signal should be obtained from only the one primary antibody PLA.) The combination of the primary antibodies is very important for the PLA detection. Different combinations can give different PLA results; even different batches of the antibody can change the results. Therefore, one should keep the same batch of the antibody when performing PLA on a number of different samples.

Since PLA can detect up to 40nm distance when two proteins are in proximity, not all the antibodies bound to the protein can establish the rolling-circle-amplification. This is also a limitation if two proteins are separated by other proteins, such as scaffold proteins, while they are still in the same complex.

It is also possible to make your own primary PLA probes. The primary antibody targeting the protein of interest can be directly conjugated with the PLA oligonucleotide probes (PLUS or MINUS), which is called single-layer PLA. Compared to the original double-layer PLA (primary antibody followed by secondary PLA probes), single-layer PLA may provide less signal, since non-specific binding is only depended on the efficiency of two primary antibodies rather than primary and secondary antibodies. However the latter combination may be able to bridge a bigger distance in the complex. In our experiments on the fetal liver tissues, we compared the single- and double-layer PLA, which showed the background in the single-layer PLA to be very low compared to the double-layer PLA; however, the real signals are much higher in

the double-layer PLA condition which may have to do with bridging the distance between the proteins (data not shown). Therefore, we recommend the double-layer PLA when primary antibodies from different species are used. If from the same species, the advantage of single-layer PLA is that even the primary antibodies from the same species can be conjugated to the oligos, one for PLUS the other one for MINUS. Importantly, if one of the primary antibodies is raised from rat, the only way to presently perform the PLA is making a single-layer PLA probe, because the commercial anti-rat PLA probe is not available anymore.

The development of PLA technique was achieved by the combination of target sequence recognition with padlock probes and proximity ligation followed by rolling-circle amplification [21]. It has been proven useful to quantify protein in liquid samples such as serum, plasma and lysates from cells or tissues. It is capable to detect very low concentrations compared to other methods. Compared to ELISA, PLA can detect a 100-fold lower protein concentration [22]. Later, the solid-phase PLA detection on cells or tissues was developed called *in situ* PLA. Recently, Ola and colleagues published a method to detect multiple protein complexes at the same time in tissue samples [23]. They have designed three different oligonucleotides which can be hybridized by distinct oligo-specific labeled detection probes. In this way, they characterized EGFR complexes that essential in the regulation of cell growth, survival and proliferation. Since GATA1 can form at least five distinct complexes to regulate its target genes in the MEL cells, it would be interesting to dissect the mechanism of these complexes during the embryonic development using the multicomplex PLA method. For instance, as mentioned in Chapter1, during the embryonic development, GATA2 is expressed earlier than GATA1, and an LDB1 complex can bind to the *Gata2* promoter. Although PLA cannot detect the real-time molecular binding events, we can still choose the different stages of ES cell differentiation to snapshot the protein-protein interactions. Another recently published article by Irene et al. describes a protocol for establishing the *in situ* detection of mRNA using padlock probes and *in situ* PLA at the same time [24]. They were able to use this new technique to study the kinetics of gene expression and the modification of a protein as a result of a signaling pathway at the same time [25]. This could be a powerful tool to dissect the complexity of GATA1/LDB1.

We tried to combine the PLA technique and fluorescent *in situ* hybridization (with Mariette van de Corput). We optimized RNA-FISH and DNA-FISH before or after performing the PLA and managed to combine DNA-FISH with PLA (**Fig d**, **Appendix**). We did the well-known interaction GATA1/LDB1 PLA in MEL non-induced and induced cells first followed by DNA-FISH. In the figure, green dots present the DNA-FISH signals, while the red dots are the PLA. We could see bright FISH signals in the cells, but the

background of the FISH was also high. Comparing the FISH signals and PLA signals, we could not clearly see an overlap of these signals in either non-induced or induced MEL cells. However, we do see some FISH and PLA signals close to each other (arrow-heads in **Fig d**, **Appendix**). The background of FISH needs to be further optimized to get clear results. For the PLA signals, we also analyzed the amount of dots per cell in the non-induced and induced MEL conditions. We found that in the non-induced condition, the GATA/LDB1 PLA signal is around 300-400 dots per cell. We found the induced cells to contain many more PLA signals (**Fig a**, **Appendix**). It is also interesting and this still needs further investigation that the PLA signals in the induced cells are a little bigger than in non-induced ones. This may be indicative of the dynamics of LDB1 complexes which can bring other long-range-distance regulatory DNA elements to the promoter region to form a chromatin hub with several LDB1 complexes.

Since PLA can detect even a single protein-protein interaction, *in situ* detection of LDB1/GATA1 and GATA1/FOG1 in the erythroblastic island in the fetal liver would provide additional evidence for their roles in erythroid development. The PLA could be performed on the fetal liver tissue using F4/80 antibody to stain the specific central macrophage. We tried to stain the erythroblastic island with F4/80 and with either CD71 or TER119 antibody on the whole fetal tissue slices. However, due to the compact tissue structure, we were not able to clearly distinguish the immature erythroid cells in the liver (data not shown). During the generation of ES-cell derived erythroid progenitor (ESEP) cells, it was easy to find structures similar to erythroblastic islands, a few erythroid cells partially attached to a single cell that was attached to the bottom of the dish. We found that these erythroblastic islands are very fragile and easily disrupted. However, Gloria et al. has developed a method to reconstruct the erythroblastic island from the bone marrow of M-CSF/GFP transgenic mice *in vitro* [26]. With this method, PLA could be used to investigate the GATA1/LDB1 and GATA1/FOG1 complexes in reconstructed erythroblastic islands. It would also be interesting to detect the LDB1 and GATA1 complexes on the embryonic tissue slices at E6.5 to E7, at which stage the primitive erythroid cells have developed in the blood islands in the yolk sac.

In summary, the study of early function of the GATA1 and LDB1 complexes using PLA technique would provide better insight of the timing of appearance and the levels of these complexes which is important for the understanding of how they regulate erythropoietic lineage development and differentiation.

References

1. Cantor, A.B. and S.H. Orkin, *Hematopoietic development: a balancing act.* Curr Opin Genet Dev, 2001. **11**(5): p. 513-9.
2. Leonard, M., et al., *Dynamics of GATA transcription factor expression during erythroid differentiation.* Blood, 1993. **82**(4): p. 1071-9.
3. Martin, D.I., et al., *Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages.* Nature, 1990. **344**(6265): p. 444-7.
4. Zon, L.I., et al., *Expression of mRNA for the GATA-binding proteins in human eosinophils and basophils: potential role in gene transcription.* Blood, 1993. **81**(12): p. 3234-41.
5. Rodriguez, P., et al., *GATA-1 forms distinct activating and repressive complexes in erythroid cells.* EMBO J, 2005. **24**(13): p. 2354-66.
6. Palstra, R.J., et al., *The beta-globin nuclear compartment in development and erythroid differentiation.* Nat Genet, 2003. **35**(2): p. 190-4.
7. Anguita, E., et al., *Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2.* EMBO J, 2004. **23**(14): p. 2841-52.
8. Meier, N., et al., *Novel binding partners of Ldb1 are required for haematopoietic development.* Development, 2006. **133**(24): p. 4913-23.
9. Mylona, A., et al., *Genome-wide analysis shows that Ldb1 controls essential hematopoietic genes/pathways in mouse early development and reveals novel players in hematopoiesis.* Blood, 2013. **121**(15): p. 2902-13.
10. Lahlil, R., et al., *SCL assembles a multifactorial complex that determines glycophorin A expression.* Mol Cell Biol, 2004. **24**(4): p. 1439-52.
11. Pevny, L., et al., *Development of hematopoietic cells lacking transcription factor GATA-1.* Development, 1995. **121**(1): p. 163-72.
12. Fujiwara, Y., et al., *Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1.* Proc Natl Acad Sci U S A, 1996. **93**(22): p. 12355-8.
13. Whyatt, D., et al., *An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells.* Nature, 2000. **406**(6795): p. 519-24.
14. Ferreira, R., et al., *Dynamic regulation of Gata factor levels is more important than their identity.* Blood, 2007. **109**(12): p. 5481-90.
15. Welch, J.J., et al., *Global regulation of erythroid gene expression by transcription factor GATA-1.* Blood, 2004. **104**(10): p. 3136-47.
16. Pope, N.J. and E.H. Bresnick, *Establishment of a cell-type-specific genetic network by the mediator complex component Med1.* Mol Cell Biol, 2013. **33**(10): p. 1938-55.
17. Jaehning, J.A., *The Paf1 complex: platform or player in RNA polymerase II transcription?* Biochim Biophys Acta, 2010. **1799**(5-6): p. 379-88.
18. Pourfarzad, F., et al., *Locus-Specific Proteomics by TChP: Targeted Chromatin Purification.* Cell Rep, 2013.
19. Woo, A.J., et al., *Role of ZBP-89 in human globin gene regulation and erythroid differentiation.* Blood, 2011. **118**(13): p. 3684-93.
20. Wu, Y., et al., *The zinc finger repressor, ZBP-89, recruits histone deacetylase 1 to repress vimentin gene expression.* Genes Cells, 2007. **12**(8): p. 905-18.
21. Landegren, U., et al., *Molecular tools for a molecular medicine: analyzing genes, transcripts and proteins using padlock and proximity probes.* J Mol Recognit, 2004. **17**(3): p. 194-7.
22. Gustafsdottir, S.M., et al., *Proximity ligation assays for sensitive and specific protein analyses.*

Anal Biochem, 2005. **345**(1): p. 2-9.
23. Leuchowius, K.J., et al., *Parallel visualization of multiple protein complexes in individual cells in tumor tissue.* Mol Cell Proteomics, 2013. **12**(6): p. 1563-71.
24. Weibrecht, I., et al., *In situ detection of individual mRNA molecules and protein complexes or post-translational modifications using padlock probes combined with the in situ proximity ligation assay.* Nat Protoc, 2013. **8**(2): p. 355-72.
25. Weibrecht, I., et al., *Simultaneous visualization of both signaling cascade activity and end-point gene expression in single cells.* PLoS One, 2011. **6**(5): p. e20148.
26. Lee, G., et al., *Targeted gene deletion demonstrates that the cell adhesion molecule ICAM-4 is critical for erythroblastic island formation.* Blood, 2006. **108**(6): p. 2064-71.

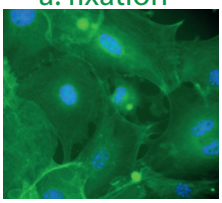
Appendix - *in situ* Proximity ligation assay (PLA)

Detection of protein interactions and protein modifications

OLINK BIOSCIENCE

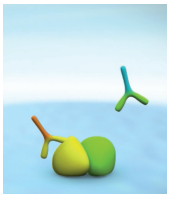
Technique principle

a. fixation



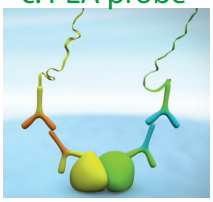
with standard IF reagents, eg. 2% PFA

b. 1st Ab



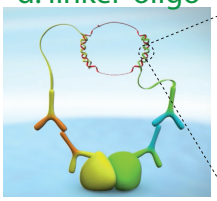
primary Ab incuation

c. PLA probe

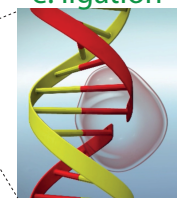


oligo-DNA conjugated Ab (PLA probe) incuation

d. linker-oligo

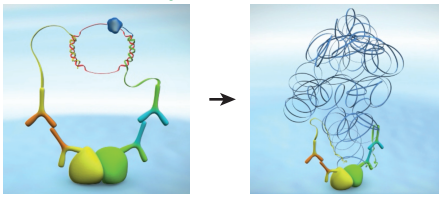


e. ligation



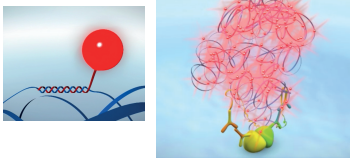
addition of oligo-DNA linker and ligation to form ring structure

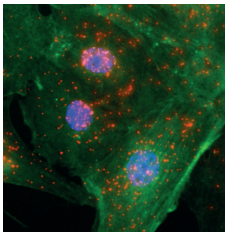
f. amplifcation (RCA)



rolling-circle-amplification (RCA) with Phi29 DNA polymerase at 37℃ for one hour, results a bundle of DNA less than 1µm

g. hybridization

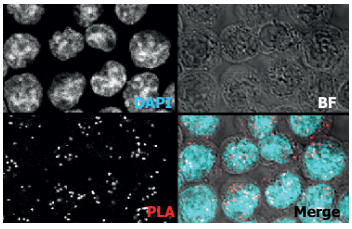


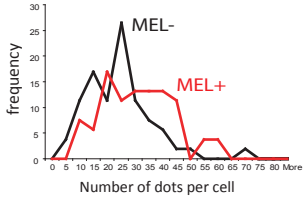


hybridization with detection probe (eg. Cy3), each interaction will present as one blob under the microscope

Applicable on different samples:

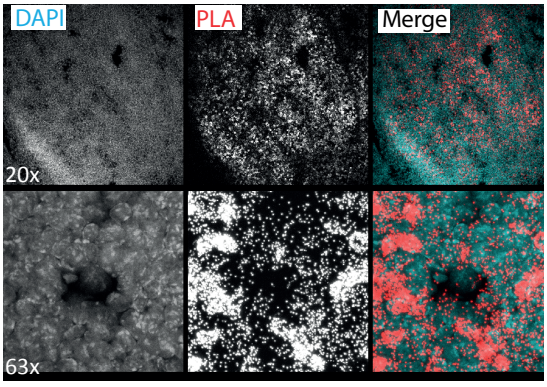
a. MEL cell line



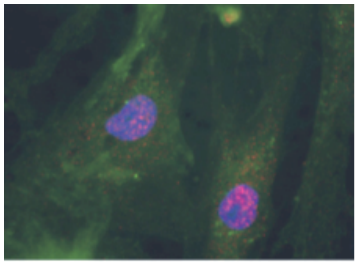


Pictures show the examples that PLA can be applied on suspension cells (a), the graph presents statistical analysis of the PLA signals per cell; the combination with immunofluorecence (b); on tissue slices (c); and the combination with fluorecent *in situ* hybridization (f). The arrows in (d) describes the FISH signals as a green dot, the arrow-heads present co-localization of PLA and FISH.

c. fetal liver tissue

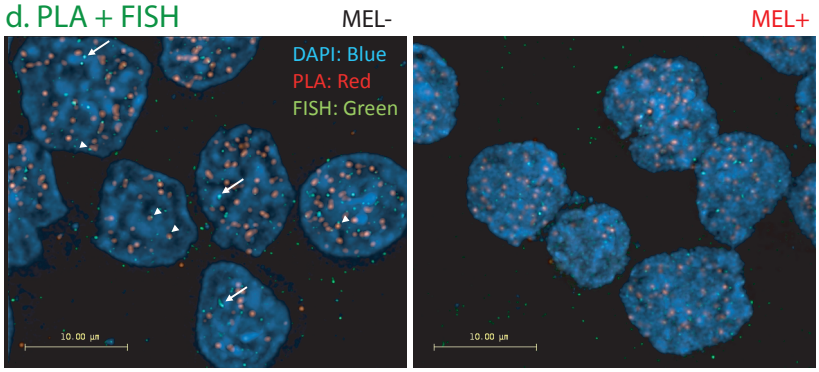


b. PLA + IF



Ola Soderberg, Methods 45 (2008) 227-232

d. PLA + FISH



Summary

Hematopoiesis is a classic model for the study of embryonic and adult stem cell differentiation. Erythropoiesis is the process of generating erythrocytes from hematopoietic stem cells (HSC).

In Chapter 1, we discuss the process of erythroid differentiation and the role of specific transcription factors in this process. Both in mouse and zebrafish, LDB1, TAL1, E2A and LMO2 have been identified as essential proteins for hematopoietic/erythroid differentiation. They can form large transcription factor complexes, including the LDB1 complex (also known as the SCL complex). This complex is important throughout erythropoiesis as it both positively and negatively modulates gene expression. The functions of LDB1 complexes have been studied in definitive erythroid differentiation; however, during embryonic development, the temporal and spatial formation of the LDB1 complex still remains unknown.

In order to answer these questions, in Chapter 2, we applied the relatively novel technology of proximity ligation assay (PLA) to detect, localize and quantify individual protein-protein interactions *in situ*, on mouse embryonic tissue slices. We show that in the fetal liver some cells specifically contain a high signal for the GATA1 and LDB1 interaction. With its DNA binding capacity, GATA1 is an important component of the LDB1 complex. However, GATA1 can also form other complexes during erythroid development such as the GATA1/FOG1 interaction. The frequency of this interaction did not increase in the cells with high GATA1/LDB1 interaction frequencies, suggesting that the increase in frequency is complex specific. We show that the GATA1 and LDB1 interaction is important in the early stage of erythropoiesis. Importantly, *Ldb1* knockdown showed that those cell populations that have high GATA1/LDB1 interaction in the PLA assay are specifically affected by LDB1 depletion. This confirms that the GATA1/LDB1 complex is essential for early erythropoiesis *in vivo*.

The regulation of gene expression is partly based on transcription factors acting on RNA polymerase II complexes. In Chapter 3, I describe interaction partners of LMO2. In LMO2 pull-down experiments, we found a specific binding of CDC73 (cell division cycle 73) which is a component of the RNA polymerase II associated factor (PAF) complexes. This factor may be one of the factors linking transcription factor complexes to the RNA polymerase II machinery.

From the results using PLA, we conclude in Chapter 4 that the GATA1/LDB1 complex, which is known to be essential for erythroid development, already emerges at day 4 of development in an *in vitro* ES cell differentiation assay. The GATA1/FOG1 complex appears only at day 5. In the fetal liver cells, we observed that GATA1 and LDB1 show a very significant interaction in the CD71 positive precursor stages of erythroid development. Finally, I discuss the PLA technique, a number of experiments that remain to be done and the future direction of the current research on LDB1 complex detection.

Samenvatting

Hematopoiese is het proces waarin een stam cel, de hematopoiese stam cel (HSC), zich ontwikkelt tot alle verschillende celtypes in het bloed. Erythropoiese is hierin het proces waarin de HSC zich ontwikkelt tot rode bloedcel, de erythrocyt. Deze processen zijn klassieke modelsystemen voor het bestuderen van stamcel differentiatie.

In hoofdstuk 1 wordt het erythropoiese proces beschreven. Daarbij wordt de rol van een aantal belangrijke transcriptie factoren behandeld. In zowel muis als zebravis is eerder aangetoond dat de transcriptie factoren LDB1, TAL1, E2A en LMO2 essentieel zijn voor een correcte erythropoiese/ hematopoiese. Deze factoren kunnen een aantal verschillende grote transcriptie factor-complexen vormen. Een zo'n complex is het LDB1 complex (ook wel SCL complex genoemd), welke belangrijk is voor een correcte erythopoiese via zowel positieve als negatieve regulatie van gen expressie. De vorming en functie van het LDB1 complex is veel bestudeerd in het definitieve erythropoiese proces. Helaas is er nog weinig bekend over de rol van het complex tijdens de ontwikkeling van primitieve erythrocyten, tijdens het embryonale stadium. Vragen als 'wanneer' en 'waar' het complex voor het eerst gevormd wordt, evenals de specifieke factoren die erbij betrokken zijn, kunnen nog niet worden beantwoord.

Om deze vragen te beantwoorden hebben we, in hoofdstuk 2, de relatief nieuwe technologie 'proximity ligation assay' (PLA) toegepast. Met deze techniek hebben we *in situ*, in weefsel coupes van muizenembryo's, individuele eiwit-eiwit interacties kunnen detecteren, lokaliseren en kwantificeren. Wij laten zien dat in de foetale lever, specifieke cellen significant meer LDB1/GATA1 interacties bevatten dan de rest van de van foetale lever cellen. GATA1 is met zijn DNA-bindingscapaciteit een belangrijk component in het LDB1 complex, maar kan tijdens de erythropoiese betrokken zijn bij een aantal andere complexen, zoals een GATA1/FOG1 complex. De frequentie van deze GATA1/FOG1 interactie is echter niet verhoogd in de cellen met een hoge LDB1/GATA1 interactie, wat aangeeft dat de toename in interactie frequentie complex-specifiek is. Wij laten zien dat de LDB1/GATA1 interactie belangrijk is in een vroeg stadium van erythropoiese. Interessant is het feit dat uit *Ldb1* depletie experimenten blijkt dat specifiek de cel populatie met een hoge LDB1/GATA1 interactie wordt beïnvloed door *Ldb1* depletie. Dit versterkt

de hypothese dat het LDB1/GATA1 complex essentieel is voor vroege erythropoiese *in vivo*.

Gen expressie wordt deels gereguleerd door transcriptie factoren die direct invloed hebben op de functie van RNA polymerase II complexen. In hoofdstuk 3, beschrijf ik de interactie partners van LMO2. In LMO2 'pull-down' experimenten hebben wij een interactie gevonden met CDC73 (cell division cycle 73), wat een component is van het RNA polymerase II-geassocieerde factoren (PAF) complex. Deze factor is dus mogelijk een connectie tussen transcriptie factor complexen en het RNA polymerase II complex.

Uit de resultaten van de PLA experimenten concluderen we in hoofdstuk 4 dat het LDB1/GATA1 complex, waarvan bekend is dat het essentieel is voor de ontwikkeling van erythrocyten, al op dag vier van een *in vitro* embryonale stamcel differentiatie wordt gevormd. Een LDB1 onafhankelijk GATA1/FOG1 interactie wordt daarentegen pas op dag 5 gevormd. In foetale levercellen hebben wij een significante LDB1/GATA1 interactie gevonden die al in CD71 positieve voorloper erythrocyten wordt gevormd. Als laatste beschrijf ik de PLA technologie, bespreek ik een aantal experimenten die nog uitgevoerd moeten worden en bespreek ik een aantal opties voor de richting waarin dit onderzoek naar LDB1 complex interacties in de toekomst verder zou kunnen gaan.

Acknowledgement / 感謝の言葉

During the last four years of my Ph.D. period, I have had the greatest privilege to work and study at Erasmus MC Rotterdam. I have formed many great memories working with these fantastic scientists and staffs. I would like to take this opportunity to express my profound gratitude and deep regards towards my colleagues, as well as show my appreciation for their support and help.

First and foremost, I would like to thank Prof. Dr. Frank Grosveld who gave me the opportunity to do my Ph.D. in his lab. During one of our first group meetings, you corrected my mistake of “Vann” into “Venn” diagram in a PPT on your computer. At that time, I strongly felt your understanding and endless kindness. In another meeting, you asked me to bring a PLA picture to count the dots in a cell by hand, in which I inherited your keen sense of precision for science. At a ski-meeting in Kleinwalsertal, you and Joost were making tools for kids and that is when, at first hand I saw you not only as an excellent scientist but also as an excellent father.

During the stressful times when I wrote my thesis, numerous times during your holiday you would take time in the middle of night to correct my insufficient English, same as Prof. Dr. Danny Huylebroeck. When I opened your files, all I could see were red lines and comment all over the pages! I will always be eternally grateful for the way you have supported me during these years. I gratefully appreciate your kindness!

Dear Dr. Charlotte, you have been my supervisor since the first day I joined the lab. I am very glad to have had the opportunity to work together and will cherish the time spent in the lab. Although after two years you left to Paris to start your new position there, yet you continued being my supervisor and helped me to solve numerous scientific problems on Skype-meetings. I will always be appreciative for your kindness. I really had so many good times with you and Eric all these years. Thank you very much!

I got a lot of help also from Marike, Jasperina and Bep on many university-related documents and thank you for helping me solve the rental problem!

I also very thankful to Prof. Dr. Joost, Dr. Niels, Prof. Dr. Sjaak and many other peoples’ care and kindness. Dr. Sjaak, thank you very much to allow me to join Kleinwalsertal meetings. The presentations and your “open discussion session” in the afternoon I enjoyed thoroughly. Prof. Dr. Danny Huylebroeck, thank you for teaching me how to take notes in meetings and also for reading my thesis in such a short notice.

What I liked the most was the international environment within in the lab. Andrea, we are the students of Dr. Charlotte and we worked on many

experiments together. I really had the best of times with you and we talked about so many other stuffs during our lunch time;) You also gave me so many great suggestions in personal life as well. I will never forget. Ralph, I am very grateful for your support in the lab for many things. I really enjoyed our language exchange although it only last for a very short period. Mary, thank you so much for helping me on my experiments. I wish you have a good time in Hubrecht. Anita, you always support me and many other lab members. Thank you very much! Petros, thank you as well for analyzing my ChIP-seq and RNA-seq data! Guillaume and Robert-Jan thanks again for discussing my experiments with me! Ruud, thank you for teaching me how to write a cover letter and CV! Mariette, thank you for helping me to combine PLA and FISH! Reinier, I appreciate you teaching me how to operate the FACS machine, without your help I would not have been able finish my experiments. Alex, thank you for organizing mice for me! Ernie, thank you for helping me for many things! Tomoko, いろいろ親切していただき、ご家族のみなさんと過ごした時間はとても楽しかったです。日本人の先輩がいることはとても嬉しかったです。また日本に行くときお邪魔します。Chiharu さんとお会いできたのもトモコさんがきっかけでした。お二人とお会いできてとても嬉しいです。In proteomics department, Dr. Jeroen, Karel, Dick, Erik-Jan, I am grateful to all of you and Karen, thank you as well and we had a great time for Olaf's defence and his dinner;) Olaf and KC, I really had a lot of fun with you guys! Ali, Alireza, Athina, Farzin, Dr. Thamar, Maria, Ileana, Jessica, Thomas, Hegias, Polynikis, Nynke, Divine, Parham, Sahar, Dr. Bob, Marta, thank you all as well! For the people that are not mentioned here, I appreciate the help of every single one of you. Everyone not only helped me achieve my potential but allowed me to grow beyond it.

During my master in Utrecht, I would like to thank Dr. Wim de Lau for supervising me in the lab of Prof. Dr. Hans Clevers. Esther, thank you so much for all the techniques you taught me. Tokameh and Robert, thanks a lot for helping me and I learnt so much from you I first arrived in the Netherlands. You opened my eyes and broadened my horizons!

I also would like to thank my Chinese colleagues and friends. 商鹏，侯珺，赵甜娜，吕鹏，非常感谢你们这些年来对我的鼓舞和帮助，使我在鹿特丹的生活充满了乐趣！我非常怀念我们曾一起到瑞士旅游。同时，我也感谢刘凡，温蓓，康宁，李宝月，徐洪，周海波，周智超，黄亚迪，栾盈，刘丽芳，赵璐，王天时，蔡蕊，任奔，杜晓燕，李天奇，吕静，马茵怡。正是有了你们的陪伴，我在荷兰的生活才变得精彩纷呈。

Looking through my life so far, I would never have been possible to stand here for my Ph.D. defence without the help and support of many people when I was studied in Japan. （私の今までの歩んできた道を振り替えてみると、日本で生活していたときに、さまざまな人のご支援、ご指導、

ご鞭撻を得たこそ、今の自分がここに立って博士論文を発表することができたのです。それらの方々のことも忘れてはなりません。）

科学知識の勉強においては、東北大学理学部生物系における学部生時代と生物研究科における修士課程においては、水野先生や大橋先生をはじめ理学部の諸先生方に大変お世話になりました。心から感謝しております。分子細胞生物学に興味を持ったのは、先生方の授業を聞いたのがきっかけでした。科学研究には無知の私は、先生方のご指導のおかげで、だんだんと専門分野に入門し、研究室の中で友達ができて大変うれしかったです。修士論文のときはかなり大変でしたが、先生方のサポートを得て、つらかった時期を乗り越えて、修士論文がうまく完成することができて、科学研究の道を辿ってきたのです。いろいろ本当に先生方と友たちの皆様にありがとうございました。

日常生活においては、最初仙台につき、東北大学に入学したときから、東北大学名誉教授の鈴木禄彌先生と奥様のハツヨ先生に大変お世話になりまして、６年間にわたって、わがままの私を自分の孫としてずっと生活上の関心から、人格の養成まで、いろいろご指導、ご教育、ご親切をいただいておりました。心から感謝しております。先生方のおかげで私は成長することができ、欠点を直すことができたのです。「蕾は長い年月を経て、栄養を十分得てから、立派な花を咲くことができるのだ。」と、今でもハツヨ先生の優しくて哲理の深い言葉が耳に響いております。大学以来、約十年間の勉強を経て、自分は一応成長してきました。いつ先生がおしゃっていた「花が咲く時期」が自分に来るのかはわからないのですが、それは段々近づいていると信じており、それを目指してこれからもがんばります。花が本当に咲いたときに、また先生にお会いに行きます。そのとき、また先生の肩を揉んであげたり、部屋のお掃除をしたりして孫としての親孝行をすると思っております。それに、東北大学病院看護学科の菊池先生には、毎回学費減免の申請にお世話になりまして、先生のお陰で学費の減免は大部分ができて、私は勉強に専念することができたので、菊池先生に本当に感謝しました。

また、大学受験の際から、学部生時代と修士時代まで、名古屋の弁護士の田代清一先生ご一家に大変お世話になりまして、心から感謝しております。いつも名古屋にお邪魔しており、先生は私を自分のお子さんの信一君と同じように親切していただき、物事のもてなし、人間との交際において、いろいろお教えいただき、大変暖かく感じておりました。仙台に居ながら、田代先生のご指導、ご支援をいただき本当にありがとうございました。

さらに、東京の弁護士の加藤満生先生と奥様の大学教授である美穂子先生にもお世話になりました。先生方のご親切、暖かい気持ちと励ましは、

私にとって日本での勉強生活の中ではかけがえのない精神的な支えのひとつになります。本当にありがとうございました。

また、東京の友達の井上善助様ご一家にお会いしたのは、二十年以上前のことでした。今まで、ずっとご親切していただいて、私は、井上さんの孫さんの夏っちゃんと愛ちゃんより、少しだけ年上でありながら、井上家の孫のように成長してきたと感じました。大学受験のさいに、夏っちゃんと藍ちゃんからくれた「祝：受験成功」との励ます絵は今でも大切に保管しております。いつも家族で楽しく過ごしたことは、私にとって小さいときからのかけがえのない心の憧れでした。みんなの励ましがあったからこそ、東京に行けば「家族」がいるんだと感じて日本で勉強する勇気があったのです。本当にありがとうございました。

それ以外にも、たくさんの方々にお会いすることができました。東北大学川内キャンパスの談話室に居られる門間さん、斉藤さんと談話室の仲間たち、それから、名古屋にいた劉先輩や今でも連絡を取れている青木君ほか、皆様のご支援、ご友情の支えで、私は日本での学業を完成して、ヨーロッパで続けて専門の科学研究をすることができたのです。皆様のご親切がなければ、今日の私の博士号授与はないといっても過言ではありません。皆さん、本当にどうもありがとうございました。また日本に行くときに、皆さんにお会いしたいと思っております。

Finally, the most important person in my life is my dear father. I would like to thank you the most. As a child you were my first teacher and as a Chinese parent, you chose a different way to educate me. I clearly remember the day when you left me at Motoyama metro station in Nagoya, when I started my studies abroad in 2002. You told me, “Be strong son. Old eagles will push young ones out of their nests. You have grown up and it is the time for you. Remember you are the son of an eagle!” I had no idea what it meant at that time. At the time I had no friends, no family around and I could not even speak the language fluently. I was in a foreign country with no idea how to cook or wash my clothes. At 10 o’clock every night I waited for the ring of the only old-style phone every night, you would chat to me for about 15 minutes then let me continue to study. Your brave, challenging and honest spirit allowed me to venture forward and aided me through numerous difficulties. This path was not an easy one, but I really learned a lot from it. You guided me to be tough and resistant and adaptable. In Japan I was able to develop my great personality, while the Netherlands broadened my horizons. I have no idea where this path follows, but this is my life and I have loved every second of it. We have been walked together all these years and now I am very glad to see you happy with Wenjuan and Yaoyao. Together we have been through tough times and you are always there for me. Thank you dad!

Sep. 2013 Xiao Yu (于 晓/暁)

Curriculum Vitae

Personal details

Name: Xiao Yu
Birth date: 18-12-1982
Birth place: Beijing, China

Education and research

2009 – 2013 **PhD**
Dept. of Biomedical Science, Erasmus Medical Center Rotterdam, the Netherlands
Lab of Prof. Dr. Frank Grosveld
Thesis title: Formation of transcription factor complexes during embryonic erythroid development

2007 – 2009 **MSc**
Dept. of Life Sciences, Tohoku University, Japan
Lab of Prof. Dr. Mizuno Kensaku
Thesis title: Roles of LIMK1 and SSH1 in cell migration in angiogenesis by VEGF signaling using FRET technique

2007-2008 Internship
Hubrecht Institute, Utrecht, the Netherlands
Lab of Prof. Dr. Hans Clevers
Project: Small molecule screening using recombinant Fc fusion proteins

2003 – 2007 **BSc**
Dept. of Biology Sciences, Tohoku University, Japan
Molecular Cell Biology
Project: Inhibitor screening for RAS and RAF interaction

Publication

Andrea Martella, **Xiao Yu**, Charlotte Andrieu-Soler, Frank Grosveld (2012). Transcriptional regulation and signaling pathways involved in the hemangioblast development. Trends in Developmental Biology. Vol.6, 53-73

Ph.D. Portofolio

Name PhD student:	Xiao Yu
Erasmus MC Department:	Biomedical Science (Cell Biology)
PhD period:	June 2009 - September 2013
Promotor:	Prof. Dr. F. G. Grosveld
Co-Promotor:	Dr. Charlotte Andrieu-Soler

General courses — Year

- Laboratory animal science (Article 9) 2010
- Epigenetic Regulation 2010
- Technology Facilities 2010
- Systems Biology Applied Bioinformatics 2010
- Next Generation Sequencing 2010
- Basic data analysis on gene expression arrays 2011
- English Biomedical Writing and Communication 2012

Workshop and conferences

- MGC PhD Workshop 2009
- EuTRACC 2nd Young Scientist Meeting, Dubrovnik Croatia ***oral presentation*** 2010
- Hemoglobin Switching meeting, Oxford, UK 2010
- 18th MGC PhD Workshop, ***oral presentation*** 2011
- Winter school, "Chromatin Changes in Differentiation and Malignancies", Kleinwalsertal, Austria, ***oral presentation*** 2011
- Monday Morning Meetings 2010-2013

Extra activities

- Committee member of organizing the PhD-Day in the Erasmus MC Rotterdam 2010
- Committee member of organizing the 18th PhD workshop in Maastricht 2011